DOCUMENT RESUME

ED 225 135 CS 006 981

AUTHOR Palincsar, Annemarie Sullivan; Brown, Ann L.
TITLE Reciprocal Teaching of Comprehension-Monitoring Activities. Technical Report No. 269.
INSTITUTION Bolt, Beranek and Newman, Inc., Cambridge, Mass.; Illinois Univ., Urbana. Center for the Study of Reading.
SPONS AGENCY National Inst. of Child Health and Human Development (NIH), Bethesda, Md.; National Inst. of Education (ED), Washington, DC.
PUB DATE Jan 83
CONTRACT 400-76-0116
GRANT NICHHD-HD-05951; NICHHD-HD-06864
NOTE 81p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Academic Aptitude; Classroom Techniques; *Cognitive Processes; Grade 7; Junior High Schools; Learning Activities; *Metacognition; Prediction; Questioning Techniques; *Reading Comprehension; Reading Difficulties; Reading Instruction; *Reading Research; Training Methods
IDENTIFIERS *Reciprocal Teaching; Summarization

ABSTRACT

Three studies were conducted to test the effectiveness of reciprocal teaching as a means of instructing seventh grade poor readers about activities they could use to increase comprehension and to ascertain that their comprehension was proceeding smoothly (comprehension monitoring). Reciprocal teaching involves having teacher and students take turns leading dialogues focusing on pertinent text features. Four comprehension-enhancing activities were emphasized: summarizing, questioning, clarifying, and predicting. The reciprocal method was compared to a traditional teaching method in the first study, with the reciprocal method producing greater gains and maintaining those gains over a longer period than the traditional method. In the second study, reciprocal teaching resulted in sizable gains on laboratory comprehension tests; reliable maintenance; generalization to classroom comprehension tests; transfer to novel laboratory tasks that tapped the trained skills of summarizing, questioning, and clarifying; and improvement of scores on standardized comprehension tests. These results were replicated in the third study in which volunteer teachers (rather than experimenters) used the method with their own reading groups. (Extensive tables of data are appended.) (FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED225135

CENTER FOR THE STUDY OF READING

Technical Report No. 269

RECIPROCAL TEACHING OF COMPREHENSION-MONITORING ACTIVITIES

Annemarie Sullivan Palincsar
and Ann L. Brown

University of Illinois at Urbana-Champaign

January 1983

University of Illinois
at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

Bolt Beranek and Newman Inc.
50 Moulton Street
Cambridge, Massachusetts 02238

The research reported herein was supported by Grants HD06864, HD05951 from the National Institute of Child Health and Human Development and from the National Institute of Education under Contract No. US-NIE-C-400-76-0116. The major part of Studies 2 and 3 were submitted as part of the requirement for the Ph.D. by the first author, who would like to thank the members of her dissertation committee for their valuable recommendations and support: Coleen Blankenship, Ann Brown, Laura Jordan, Steve Lilly, and David Pearson.

CS806981

Reciprocal Teaching of Comprehension-Monitoring Activities

One of the most powerful tools of the discipline of cognitive engineering or applied cognitive science (Norman, 1980) is the training study (Chipman, Segal, & Glaser, in press). Guided by emergent theoretical analyses of the processes involved in a particular academic domain, researchers have designed cognitive skills training studies that have resulted in significant improvement in such areas as physics and mathematics problem solving (Larkin, Heller, & Greeno, 1980), writing (Bereiter & Scardamalia, 1980) and various of the multifaceted skills that underlie reading and studying (Brown, Campione, & Day, 1981; Brown, Palincsar, & Armbruster, in press).

In order to engineer significant improvement in academic skills, however, the researcher needs a detailed specification of the processes underlying adequate performance, and a correspondingly detailed task analysis of an instructionally relevant activity (Resnick & Glaser, 1976). In addition, adequate diagnosis of the student is required so that the level of instruction can be calibrated to the starting competence of the learner (Brown, Bransford, Ferrara, & Campione, in press; Klahr & Siegler, 1978). Finally, clear criteria of success should include factors such as the interpretability, reliability, durability, and transferability of any effects of the intervention (Brown & Campione, 1981). While these guidelines are widely accepted, to the best of our knowledge, they have never been incorporated within a single program of research. In this paper we will report a series of studies that, as a package, does include all these factors.

The object of our training studies was to enhance reading comprehension and comprehension-monitoring; i.e., to instruct students regarding activities they could engage in both to promote understanding and to ascertain that comprehension is proceeding smoothly. Although we are still far from a detailed task analysis of reading comprehension, there are several overlapping skills that have been repeatedly mentioned as prime comprehension-fostering activities in a variety of recent theoretical treatments (cf. Baker & Brown, in press a,b; Brown, 1980; Collins & Smith, in press; Dansereau, 1980; Markman, in press). These activities include: (a) clarifying the purposes of reading, i.e., understanding the task demands, both explicit and implicit; (b) activating relevant background knowledge; (c) allocating attention so that concentration can be focused on the major content at the expense of trivia; (d) critical evaluation of content for internal consistency, and compatibility with prior knowledge and common sense; (e) monitoring ongoing activities to see if comprehension is occurring, by engaging in such activities as periodic review and self-interrogation; and (f) drawing and testing inferences of many kinds, including interpretations, predictions and conclusions.

In this series of studies, we concentrated on four, commonly accepted, comprehension-enhancing activities: summarizing, questioning, clarifying and predicting. All of these activities appear as academic tasks in their own right; for example, it is a common practice to call on a student to summarize or answer questions on a passage. But, in addition, these activities, if engaged in while reading, serve to enhance comprehension and

afford an opportunity for the student to check whether it is occurring. That is, they can be both comprehension-fostering and comprehension-monitoring activities if properly used. Self-directed summarization is an excellent comprehension-monitoring technique (Brown & Day, in press; Brown, Day, & Jones, in press; Day, 1980; Linden & Wittrock, 1981). Monitoring one's progress while reading, to test whether one can pinpoint and retain important material, provides a check that comprehension is progressing smoothly. If the reader cannot produce an adequate synopsis of what she is reading, this is a clear sign that comprehension is not proceeding smoothly and that remedial action is called for.

Similarly, self-directed questioning concerning the meaning of text content leads students to a more active monitoring of their own comprehension (André & Anderson, 1978-79). Thus, closing one's eyes (metaphorically) and attempting to state the gist of what one has read, and asking questions of an interpretive and predictive nature (Collins & Smith, in press) are activities that both improve comprehension and permit students to monitor their own understanding. These are also the kinds of active and aggressive interactions with texts that poor readers do not engage in readily; the need for explicit instruction in comprehension-enhancing activities is particularly acute in the slow-learning student (Brown & Palincsar, 1982).

One of the primary problems facing those who would design cognitive skills training is deciding what level of help students need. Discussions of this point have centered around the issue of specific and general

skills. To illustrate this problem Newell (1979) introduced the metaphor of an inverted cone of skills. At the bottom of the cone, the broad base, he conceived of a large set of specific powerful routines that are applicable to a limited number of domains; they are powerful in that once they are accessed, problem solution should follow (assuming only that they are executed properly). As we move up the cone, there is a tradeoff between generality and power. At the tip of the cone, there are a few highly general but weak routines--general in that they are applicable to almost any problem-solving situation but weak in that they alone will not lead to problem solution. Examples here include exhortations to stay on task or to monitor progress. These are weak in that, for example, merely noticing that progress is not being made or that learning is not occurring cannot rectify the situation unless the student brings to bear more powerful routines that can result in better learning.

A great deal of the existing training research has concentrated on either a subset of very specific skills or on the very general "metacognitive" level (see Brown et al., in press; Brown & Palincsar, 1982). However, rather than teaching a large number of specific routines or some extremely general supervisory ones, an alternative approach would be to identify and teach packages of skills (Campione & Armbruster, in press; Dansereau, Collins, McDonald, Halley, Garland, Diekkoff, & Evans, 1979) that include the coordination of both. An excellent example of such an approach comes from the "self-instruction" work inspired by cognitive behavior modification techniques (Meichenbaum, 1977; Meichenbaum & Goodman,

1971). Initial work in this vein could be characterized as concentrating on the weak general methods. Typically, the student is trained in general coping skills such as "slow down," "look carefully at all your choices," "check your work," etc. In general, these programs produce excellent short-term results with children who have at their disposal the necessary task-specific skills, and whose learning problems reside primarily in controlling and overseeing the use of those skills. Hyperactive, impulsive children respond very well to such regimes.

These self-control programs are, however, insufficient for problem learners who do not already know how to perform the task specific elements of the problem. To deal with this eventuality, researchers in cognitive behavior modification have added to the general coping litany direct instruction in task specific elements; this is termed response guidance. To illustrate, Meichenbaum reports significant improvement on standardized reading tests by junior high school students trained to use the following self-directions:

> Well, I've learned three things to keep in mind before I read a story and while I read it. One is to ask myself what the main idea of the story is. What is the story about? A second is to learn important details of the story as I go along. The order of the main events or their sequence is an especially important detail. A third is to get to know how the characters feel and why. So, get the main idea, watch sequences, and learn how the characters feel and why. (Meichenbaum & Asarnow, 1978, p. 17)

We have argued elsewhere (Brown, 1978; Brown & Campione, 1978; Brown & Palincsar, 1982) that ideal cognitive skills training programs should include practice in the specific task appropriate strategies (skills training), explicit instruction in the orchestration, overseeing and

monitoring of these skills (self-regulation training) and information concerning the significance of those activities and their range of utility (awareness training). Therefore, the training vehicle we chose was a combined package involving all three levels of instruction. The "skills" trained were summarizing, questioning, predicting, and clarifying. The students received explicit instruction, extensive modeling, and repeated practice in concrete versions of these activities. Second, the students were constantly reminded to engage in these activities while reading, indeed to read for the purpose of performing these activities for themselves. They were instructed not to proceed until they could summarize, clarify, and answer questions on each segment of text (self-regulation training). And, third, the students were constantly reminded that these activities were to help them improve and monitor their own comprehension, shown that their performance improved dramatically when they did so, and told that they should always engage in them while reading for academic purposes (awareness training).

We embedded these activities within a promising training procedure that permits explicit modeling of such comprehension-fostering activities, that of reciprocal questioning. Manzo (1968) introduced a variant of this with his ReQuest procedure. Teachers and small groups of remedial-reading students took turns asking themselves questions about what they were reading. Questions followed every sentence, a procedure that would not encourage synthesis across larger segments of text. And the types of questions modeled and generated were not necessarily optimal. For example,

one teacher modeled the question "what was the third word in the first sentence?" Even so, Manzo reported significant improvement in standardized reading comprehension scores.

Frase and Schwartz (1975) also had students taking turns generating or answering questions. Regardless of which role the students assumed, they performed better than when engaged in silent reading. Even though training was not extensive, and again there was no attempt to ensure adequate quality of questions, the intervention produced a modest but reliable effect. Given these promising precursors, we decided to adopt the reciprocal teaching method where, in addition to question generating, we added the activities of reciprocal paraphrasing, clarifying, and predicting.

Three studies are reported. In the first, a comparison was made between two interventions: locating information, where practice was provided in using the text intelligently in order to answer explicit and inferential questions, and the reciprocal teaching intervention. Only the reciprocal teaching procedure was featured in Studies 2 and 3. The studies also varied in how closely they approximated the kinds of teaching settings that could feasibly occur in schools. In the first study, individual teaching was conducted and in the second the investigator worked with groups of two students, again on a "pull-out" basis. In the third study, volunteer reading teachers attempted to implement the intervention in their existing reading groups. In all studies, training was aimed at students diagnosed as particularly in need of assistance with reading comprehension,

i.e., those students who decode adequately but for a variety of reasons comprehend poorly.

Finally, in these studies, we attempted to address a blatant gap in the cognitive skills training literature in that multiple and stringent criteria of success were examined. These included (a) reliable improvement on the training task; (b) independent evidence of improvement in the strategies trained; (c) improvement in the students' independent reading of novel passages; (d) durability of the effect of training; (e) generalization of the effects across settings, notably to the classroom; and (f) transfer to novel tasks that demand the same underlying processes but differ in surface structure from the training vehicle. Previous studies have included no more than two of these; indeed, the majority have included only the first. We were particularly concerned with including transfer measures to assess what, if any, general skills are improved by the intervention.

STUDY 1

Method

Subjects. Four seventh grade students were selected from a middle school located in a midwestern city of approximately 60,000 residents. A team of five seventh grade teachers were asked to nominate students whom they suspected to be adequate decoders but poor comprehenders. The teachers named 13 out of a possible 113 seventh grade students. The 13 nominees were then tested to determine if, indeed, their decoding was adequate. Their oral reading rates were found to meet a criterion of at

least 80 wpm correct with two or fewer errors, when reading grade appropriate materials. This criterion was established by Lovitt and Hansen (1976) as the minimum acceptable decoding fluency for instructional purposes. Next their putative comprehension difficulties were determined by establishing that the subjects met three criteria: (a) standardized reading comprehension scores at least two years below grade level; (b) scores below the 20th percentile on reading comprehension tests administered to the entire seventh grade population in their school; and (c) baseline performance of below 20% correct on the experimental task.

Seven of the 13 children nominated were determined eligible to participate in the study. One student elected not to participate. Four of the remaining six were selected at random to participate.

These four students ranged in age from twelve years, three months to thirteen years, five months. They were not diagnosed as learning disabled, but three of the students had Metropolitan reading comprehension scores that were three years behind grade level, and one was two years delayed. In addition, three of the students had WISC IQ scores (administered by the authors) in the low normal range (74, 89, 89) while one was average (108). Three of the four students were in the seventh percentile for reading comprehension compared with their age mates; one was in the fifteenth percentile (the outlying score in all cases was not generated by the same child). Two of the students were male, and three were black.

Materials.[1] A total of 102 400-word passages of an expository nature were employed during the study. Sources for these passages included: The

Headway Program (Open Court Publishing Co., 1979), Nature at its Strangest: True Stories from the Files of the Smithsonian Institution's Center for Short-Lived Phenomena (Cornell, 1974), Reader's Digest Skill Builders Series (Reader's Digest Services, Inc., 1977). The readability of passages selected was assessed with the Fry Readability Formula (Fry, 1977) to determine that passages fell within a seventh grade readability range. Ten comprehension questions were constructed for each passage. Using the classification scheme developed by Pearson and Johnson (1978), the questions were representative of two types--text explicit and text implicit. As the name suggests, the answers to text explicit questions appear explicitly in the text. The answers to text implicit questions are also in the text, but they require the reader to integrate information across sentences or paragraphs.

Six 400-word passages used in the social studies class probes were taken from the text currently in use in the seventh grade program, In a Race with Time: An Introduction to Latin America (Macmillan Co., 1972). Ten comprehension questions were constructed by the first author, again including text explicit and text implicit queries.

Procedures

General procedure. Each day of the study the students were presented with a 400-word assessment passage which they were required to read silently in order to answer ten comprehension questions from memory after completing the passage. The students were told to ask for assistance with any words they could not read or understand. Upon completing the passage,

the students were given ten comprehension questions which were asked and answered orally. It is the responses to these assessment passages that are reported as data throughout. After responding to the comprehension questions, the students were asked to read the passage aloud to determine correct and incorrect reading rates. This permitted an evaluation of the extent to which decoding might have impeded comprehension.

Baseline. During baseline, administration of the assessment passage was the only activity to occur.

Intervention. During the intervention phases of the study, the assessment passage was preceded by a training passage on which the investigator and student interacted. There were two forms of intervention: (a) locating information and (b) reciprocal teaching.

During locating information, the students were asked to read a passage silently and carefully in order to answer comprehension questions. They were reminded to ask for assistance with any word that they could not read or understand. Upon completing the passage, the students were asked ten comprehension questions. The investigator praised correct responses. Corrective feedback was provided for incorrect responses by guiding the student back into the passage to the appropriate paragraph where the answer could be found. If necessary, the line(s) where the answer could be found was given, as well as prompts to help the students find the answer. During the procedure, the students were being taught that the answers to the questions could be found with a little work with the text; a proposition that they greeted with some surprise! Answers to questions were discussed and mutually agreed upon by student and investigator.

Prior to initiating the reciprocal teaching intervention, the students were told about the four activities they would engage in: summarizing, questioning, predicting, and clarifying. During the intervention, the investigator and the student engaged in an interactive learning game that involved taking turns in leading a dialogue concerning each segment of text. If the passage were new, the investigator called the student's attention to the title, asked for predictions based upon the title, and discussed the relationship of the passage to prior knowledge. For example, if the passage were entitled Ship of the Desert, the investigator and student would speculate what the passage might concern and would review what they knew about the characteristics of the desert. If the passage were partially completed, the investigator asked the student to recall and state the topic of the text and several important points already covered in the passage.

The investigator then assigned a segment of the passage to be read (usually a paragraph) and either indicated that she would be the teacher or assigned the student to teach that segment. The investigator and student then read the assigned segment silently. The teacher for that segment proceeded to first ask a question, then summarize, and offer a prediction and clarification when appropriate.

The adult teacher provided the guidance necessary for the student teacher to complete the preceding activities through a variety of techniques: prompting, "What question did you think a teacher might ask?"; instruction, "Remember, a summary is a shortened version, it doesn't

include detail"; and modifying the activity, "If you're having a hard time thinking of a question, why don't you summarize first?"

The adult teacher also provided praise and feedback specific to the student's participation: "You asked that question well; it was very clear what information you wanted"; "Excellent prediction, let's see if you're right"; "That was interesting information. It was information that I would call detail in the passage. Can you find the most important information?" After this type of feedback, the adult teacher modeled any activity which continued to need improvement: "A question I would have asked would be . . . "; "I would summarize by saying . . . "; "Did you find this statement unclear?"

After proceeding through the passage in this manner for a period of 25 to 30 minutes, the assessment procedure began. All dialogues were tape recorded so that qualitative changes could be assessed.

Throughout the interventions, the students were explicitly told that these activities were general strategies to help them understand better as they read, and that they should try to do something like this when they read silently. It was pointed out that being able to say in your own words what one has just read, and being able to guess what the questions will be on a text, are sure ways of testing oneself to see if one has understood.

Maintenance. The maintenance phase began immediately after the last day of intervention and was conducted in the same manner as baseline. Students silently read the assessment passage, completed the accompanying questions, and then read the passage orally.

Generalization probes. During the various phases of the intervention, probes were taken to determine if gains demonstrated in the experimental setting would be seen in the classroom. These probes were conducted during the social studies period by the social studies teacher. The students read passages from their social studies texts and then answered ten comprehension questions (in writing). No hint was given that these exercises were in any way related to the experimental procedure.

Longterm follow-up. Six months after the last intervention day, the students were retested for eight additional days, four of maintenance, followed by two days where the reciprocal teaching was reintroduced, followed again by two maintenance days. Again, all new passages were used in the six month follow-up and only the data from the independently read assessment passages will be reported.

Design

The study employed a multiple baseline across two randomly assigned groups of two students. A cross-over design format was used for the two interventions to control for possible order effects. In addition, maintenance phases were introduced between interventions. Group 1 experienced the following sequence of phases: baseline (6-8 days), locating information (10 days), maintenance 1 (6 days), reciprocal teaching 1 (10 days), maintenance 2 (6 days), reciprocal teaching 2 (3 days). Group 2 experienced the same sequence with the exception that the order of locating information and reciprocal teaching was reversed. The institution of reciprocal teaching 2 as the last phase for each student (with the

exception of Student 2) was not initially planned. However, when it was observed that students who had locating information prior to reciprocal teaching appeared to profit more from the intervention, the decision was made to return all students to reciprocal teaching. This was possible for all students except Student 2, who was not available due to the Christmas recess. In summary, during all phases of the study, the students read assessment passages and answered ten questions, and this is all they did during baseline and maintenance phases. During the intervention phases, whether locating information or taking part in reciprocal teaching, the training took place prior to the assessment passages and employed a different training text. All data to be reported were gathered from the independently-read, daily assessment passages.

Results and Discussion

Before reporting the major data base, the percent correct each day on the assessment passages, two points should be mentioned. First, oral reading (decoding) accuracy was assessed on a daily basis. Correct and incorrect reading rates were stable throughout the study and suggest that decoding was not an impediment to comprehension. The mean correct rate for the four students ranged from 101 to 123 words per minute. The mean error rates for the four students ranged from .9 to 1.7 words per minute. These data will not be discussed further.

In addition, examination of responses to text explicit as opposed to text implicit questions revealed that students did not perform differentially on these question types. Therefore, the results reported

and discussed are percentage correct on total comprehension measures, i.e., ten questions daily.

Daily Comprehension Measure

The percent correct each day on the assessment passages is plotted in Figure 1. The most successful intervention was the sequence where

---

INSERT FIGURE 1 ABOUT HERE

---

locating information was followed by reciprocal teaching, the intervention given to Group 1. Although performance was variable, a gradual improvement across days was found. Performance increased from approximately 15% correct during baseline to 50% correct in the locating information phase, and the students maintained this level of performance, although Student 2's performance was quite variable. When the reciprocal teaching was introduced, both students achieved their most accurate and stable performance when comparing these results to previous phases. Mean accuracy for both students was 80%, a level that was also maintained during the maintenance phase for both students and during the brief re-introduction to the reciprocal teaching for Student 1. Remember that these scores, shown in detail in Figure 1, were obtained on the privately read assessment passages, i.e., different texts that the students read independently after their interaction with the instructor. What was learned during the instructional sequence was used independently by the learners.

In contrast, the performance of students in Group 2 was not quite as impressive (see Figure 1). They received the main reciprocal teaching sessions before the locating information instruction, and while performance did improve (from 15% at baseline to 50% correct during intervention and maintenance), it never reached the level set by the Group 1 students. The introduction of locating information resulted in a decrease in Student 3's accuracy and more variability in day-to-day performance than observed in previous phases. Student 4's performance during locating information is characterized by a decelerating trend line. For this reason, at the end of the last maintenance phase, the Group 2 students were reintroduced to the reciprocal teaching procedure. This resulted in the most accurate performance for any phase for both students. Student 3 averaged 77% accuracy while Student 4 averaged 87% accuracy. Apparently, the most appropriate order of these treatments is corrective feedback followed by strategy training.

A series of analysis involving planned comparisons were conducted to examine various facets of the data. The analyses supported the visual observations made previously that performance during the first intervention improved significantly over baseline performance for both reciprocal teaching, $F(1,24) = 68.23$, $p < .0001$, and locating information, $F(1,24) = 85.55$, $p < .0001$. However, performance during the maintenance phase which followed reciprocal teaching was significantly higher than the maintenance phase which followed locating information, $F(1,24) = 9.40$, $p < .005$. Performance during the reintroduction to the reciprocal teaching was

significantly greater than performance during the first reciprocal teaching phase, $F(1,24) = 25.45$, $p < .0001$.

Longterm follow-up. The results of the daily comprehension measures taken six months after the end of intervention are shown in Figure 2.

INSERT FIGURE 2 ABOUT HERE

Figure 2 also includes the results of original baseline and reciprocal teaching phases for all subjects. The six month intervention included four days of unprompted maintenance and performance declined significantly from the level set at the last intervention, $F(1,24) = 29.17$, $p < .0001$. Note, however, that the students averaged 60% correct on the long-term maintenance, a sizable improvement over their baseline performance of 15%. Note also that after only two days of renewed reciprocal teaching, performance for Students 1 and 3 returned to an 80% level and for Students 2 and 4 returned to 90% correct, suggesting a sizable savings. This increase was reliable, $F(1,24) = 33.49$, $p < .0001$ and was maintained as there is no significant difference between Maintenance 4 and Reciprocal Teaching 3.

Generalization to classroom settings. Throughout the study a series of five probes was made in the social studies classroom setting to see if the students would show any improvement on the identical task of answering ten comprehension questions on a text. All students began the study below the 15th percentile on this task compared with the remaining seventh graders in

their school. The changes in percentile ranking as a function of intervention phase are shown in Table 1. Performance fluctuated widely which was not surprising in light of the fact that little was done to promote generalization to the classroom, e.g., the classroom teaching did not encourage the use of strategies and the students received no feedback regarding classroom performance. However, the following mean gains in percentile ranks were obtained between the baseline and final probes: Student 1 = 20, Student 2 = 46, Student 3 = 4, and Student 4 = 34.

INSERT TABLE 1 ABOUT HERE

Qualitative Changes in Dialogue[2]

In addition to the improvement in number of questions correct on the independently read texts, there were several qualitative indices of improvement due to training. For example, although the students were repeatedly encouraged to ask for help with any word(s) they had difficulty reading or understanding, until the reciprocal teaching intervention was introduced, not a single student requested this type of assistance. The fact that students did request this help during reciprocal teaching, combined with the fact that students were also beginning to re-read might serve as further testimony that they were more actively monitoring their comprehension in the reciprocal teaching condition.

During the reciprocal teaching sessions the students took turns leading the dialogue, trading places with the experimenter. Initially, the experimenter modeled appropriate activities but the students had great difficulty assuming the role of dialogue leader when their turn came. The experimenter was forced to resort to constructing paraphrases and questions for the students to mimic. In this initial phase, the experimenter was modeling effective comprehension monitoring strategies but the student was a relatively passive observer.

In the intermediate phase, the students became much more capable of playing their role as dialogue leader and by the end of ten sessions were providing paraphrases and questions of some sophistication. For example, in the initial sessions, 46% of questions produced by the students were judged as non-questions or as needing clarification. By the end of the sessions only 2% of responses were judged as either needing clarification or non-questions. This improvement in questioning is shown in Figure 3.

INSERT FIGURE 3 ABOUT HERE

Unclear questions drop out and are replaced over time with questions focusing on the main idea of the segment of training. Comparing the proportions of questions that were categorized as generated with assistance or unclear, there were significantly fewer at the conclusion of treatment than at the beginning, $z = 6.40$, $p < .0001$. In contrast, there were significantly more main idea questions generated during the final segment

of training than during the first segment, $z = 4.73$, $p < .0001$. Examples of questions needing clarification, main idea and detail are shown in Table 2.

INSERT TABLE 2 ABOUT HERE

A similar improvement in summary statements was found and these data are plotted in Figure 4. When evaluating the proportions of summary statements generated by the students throughout the course of reciprocal teaching, there were significantly fewer incomplete/incorrect statements, $z = 3.89$, $p < .0001$ and detail statements, $z = 3.13$, $p < .0001$ during the final third of training as compared to the first third. Concurrently,

---

INSERT FIGURE 4 ABOUT HERE

---

there was a significant increase in main idea summary statements over time, $z = 4.84$, $p < .0001$. Examples of these summary statements are shown in Table 3. With repeated interaction with a model performing appropriate

---

INSERT TABLE 3 ABOUT HERE

---

questioning and paraphrasing activities, the students became able to perform these functions on their own. Over time the students' questions

became more like the model's, being classified as inventions, i.e., questions and summaries of gist in one's own words, rather than selections, repetitions of words actually occurring in the text (Brown & Day, in press). For example, an early occurring form of question would be to take verbatim from the text "plans are being made to use nuclear power" and append the question with the inflection "for what?" Later forms of questioning were more likely to be paraphrases of the gist in the students' own words. For example, reading a passage about fossils, one student posed the following question: "When an animal dies, certain parts decay, but what parts are saved?" This question was constructed by integrating information presented across several sentences.

In summary, students in Study 1 showed a dramatic improvement in their ability to answer comprehension questions on independently read texts. This improvement was durable in the resource room setting and showed some tendency to generalize to the classroom setting. In addition, qualitative improvement in the students' dialogues reflected their increasing tendency to concentrate on questions and summaries of the main idea. The reciprocal teaching procedure was a powerful intervention for improving comprehension. While locating information was a simpler procedure to implement and is certainly superior to no intervention, the students found locating information to be somewhat aversive, the effects were not as impressive nor as enduring as the effects of reciprocal teaching. In addition, while locating information may suggest to students the need to slow down or read more carefully, it does not facilitate the explicit instruction of skills

which students might actively engage while reading.

STUDY 2

Encouraged by the success of the initial study, we decided to replicate the main features of the successful reciprocal teaching procedure with six additional students, in three groups of two. In addition to group size, the second study also differed from the first in that (a) only the reciprocal teaching training was given; (b) a criterion level of 70% correct on four out of five consecutive days was established; (c) students received explicit (graphed) knowledge of results; and (d) tests of transfer were included.

The tests of transfer were selected because we believed that they tapped the skills taught during the reciprocal teaching, and, pragmatically, because a considerable body of prior work had established "normal" levels of performance for seventh graders. Two of the four transfer tests were measures of the two most frequently engaged in activities during the reciprocal teaching sessions, summarizing (Brown & Day, in press) and predicting questions that might be asked concerning each segment of text (Wong & Jones, 1981). In addition, two other tests were used as measures of general comprehension monitoring, error detection (Markman, 1978; Harris, Kruithof, Terwogt, & Visser, 1981) and rating importance of segments of narratives (Brown & Smiley, 1977).

Method

Subjects. The six students that took part in the study were selected from 41 teacher-nominated students in the developmental reading classes of a middle school of a middle-size midwestern city with a population of 100,000. The 41 candidates for inclusion were screened to see if they met the decoding and comprehension criteria described in Study 1. Of the 41 students, 29 met the criteria that determined the diagnosis of poor comprehender, but only 16 met the decoding criteria that placed them at grade level. Six students were selected randomly from these 16 to participate in the training. Six other eligible candidates were administered a sample of the baseline and follow-up passages with four weeks (the length of the reciprocal teaching phase) intervening. The six were also administered the pre- and post-test transfer measures, again with four weeks intervening. In addition, 13 seventh graders with no reading or other academic problems took baseline assessments and all the transfer measures.

One of the six experimental students was male, all but one were white. Details of their standardized scores are shown in Table 4. The students

---

INSERT TABLE 4 ABOUT HERE

---

were of low average IQ, their decoding fluency on seventh grade texts was at or above 100 wpm, with approximately one error per minute. The students on the average were two and a half years delayed on standardized scores of

reading comprehension.

Materials. Lengthier materials than were used in Study 1 were selected for Studies 2 and 3 to allow for more opportunity for student participation in the group settings. A total of 13 passages were available for training, averaging 1500 words in length. They were selected from the following reading series: Reading Unlimited (Scott Foresman, 1976); Keys to Reading (The Economy Company, 1980); Adventures for Readers (Harcourt Brace Jovanovich, 1979); Reading 720 (Ginn and Company, 1976); Corrective Reading Decoding (Science Research Associates Inc., 1978); Serendipity (Houghton-Mifflin, 1974). All of the passages were expository and represented a range of topics including: poisonous snakes, solar energy, the Inca civilization, lightning, and carnivorous plants. The passages were selected after determining that they obtained a seventh grade readability according to the Fry Readability Formula. The passages were sequenced in an easy to hard series so that the first three passages lent themselves more readily to determining the main idea and extracting questions, i.e., the paragraphs were relatively short and there was frequent use of topic headings.

In addition to the training passages there were a total of 35 assessment passages with their accompanying sets of ten comprehension questions. The assessment passages were expository, written at a seventh grade readability level (according to the Fry formula), ranged in length from 400-475 words and were taken from the same reading programs cited for the training materials.

The ten comprehension questions per passage were constructed by the first author using the Pearson and Johnson (1978) classification of: (a) text explicit, where the answer was explicitly mentioned in the text; (b) text implicit, where the answer needed to be inferred by combining across adjacent segments of text; and (c) script implicit, where the answer must be arrived at by considering text in relation to prior knowledge concerning the topic in question. Independent raters agreed to the classification of question types for each passage and to the fact that the questions were of approximately equal difficulty across passages (see Palincsar, 1982, for full details).

Procedures

There were four phases to the study. As in Study 1, each student was given a daily assessment passage on which she answered ten comprehension questions and this was all that occurred on baseline and maintenance days. On intervention days the assessment passage was preceded by the reciprocal teaching intervention, identical to that described in Study 1. The phases of Study 2 were as follows: (a) variable baseline consisting of four days for Group 1, six days for Group 2, and eight days for Group 3; (b) reciprocal teaching intervention consisting of approximately 20 days; (c) maintenance consisting of five days of testing at the termination of training; and (d) longterm follow-up that took place eight weeks later (3 days). All students were appraised of their progress on a daily basis. They were shown graphs depicting the percentage correct for the previous day's assessment.

Generalization probes in the classroom were taken five times during the course of the study. The probes occurred in both the social studies and science classes, with no notification given to the students that these tests were part of the study. The entire class took the tests as part of the regular work in the class. The passages used in the generalization probes were also taken from the books actually in use in the classes (World Geography, Follett Social Studies, Follett Publishing Co., Chicago, 1980; and Life Science, Silver Burdett Co., Glenview, IL, 1979). The text segments were selected in consultation with the regular classroom teachers and featured material not yet introduced to the students, so that reading comprehension was not confounded with prior instruction. Each segment was approximately 450 words long and written at a seventh grade readability level (Fry, 1977). The ten comprehension questions met the same criteria as those described for the daily comprehension questions. The teachers approved the questions as "the type one should ask students" although they expressed considerable doubt that these particular students (poor comprehenders) could handle them.

Transfer tests were included in a pre- post-test format. Prior to and on termination of the study, the students were given four tasks, summarizing, predicting questions a teacher might ask, error detecting and rating importance level.

Summarizing main ideas. The procedure here was a simplified version of that used by Brown and Day (in press) to examine students' use of various macrorules (Kintsch & vanDijk, 1978) for condensing texts. The

five principal rules were: (i) deletion of trivia; (ii) deletion of redundancy; (iii) superordination, where a list of exemplars was replaced with a superordinate term; (iv) selection of a topic sentence to serve as a scaffolding of the summary and (v) invention of a topic sentence for a paragraph where one was not explicitly stated (see Brown & Day, in press, for details). As previously found by Brown and Day, unaided seventh graders had great difficulty using any but the deletion rules. Therefore, we simplified the task and 1) told students what the five rules were and wrote them on the board with examples and 2) asked students to apply these rules to the two written texts in front of them (texts selected from Brown & Day, in press, and Day, 1980). We did not demand the third stage in the Brown and Day studies, i.e., that the students should actually write their summary after preparing the summary texts.

Question prediction. The ability to generate important and clear questions was a skill which received considerable focus during training. The following measure was included to assess the accuracy with which students could identify and construct "teacher-like" questions. The students were given four randomly assigned passages, two prior and two following the study. They were asked to predict and write ten questions a classroom teacher might ask if testing the student's knowledge of the passage. The passages were taken from material written at a seventh grade level (Fry, 1977).

Detecting incongruities. One popular index of comprehension monitoring is the ability to detect errors or anomalies in text (Baker &

Anderson, 1981; Garner, 1980; Harris et al., 1981; Markman, 1977, 1979). The Harris et al. task (1981) was selected for inclusion in the pre- post-measures. Harris et al. presented students stories line-by-line. Each story contained nine lines, one of which was anomalous to the title of the story. Harris et al. improved upon traditional measures of comprehension monitoring by recording the time students spent reading each line as well as overt indications that the students detected the anomalous passage.

There were four stories used in the Harris et al. study. To increase the number of measures, we constructed two more stories using the same guidelines detailed by Harris et al. (1981). For the purpose of including one story in both the pre- and post-testing which contained no incongruities, two stories were constructed using an identical format but containing no errors. The eight stories were randomly selected such that each student was presented two of the Harris stories, one of the newly constructed anomalous stories, and one of the errorless stories during pre- and post-testing.

Each story was presented, line-by-line, on an Apple II computer. The students were told to read each line and say "yes" if the line made sense in the story or "no" if the line didn't make sense. After reading and appraising each line of the story, the entire story was returned to the screen. To avoid contriving a situation in which students began to have expectations regarding the passages, a very general probe was used. For any story evaluated by the student as all correct, the examiner said, "Here is the entire story. You have decided that every line made sense. Is that

correct?" For any line evaluated by the student as incorrect the investigator pointed to the line(s) and said, "Can you tell me why this line doesn't make sense in the story?"

Rating of thematic importance. Four passages prepared for the Brown and Smiley studies (1977, 1978) and selected as measures of sensitivity to main idea and detail information were randomly administered to each student. Two were administered before the intervention and two after the intervention. The students were asked to read first the intact stories. They were then presented the stories with each idea unit typed on a separate line. The students were told that the stories were to be rewritten for the purpose of fitting them into tiny doll house books and that they were to choose only the most important lines. It was explained that they were to first delete N lines (1/4 of the text) by crossing out the least important with a blue pencil. They were then asked to eliminate the next N unnecessary lines using a green pencil. Finally, they were asked to cross out another N lines in red leaving only the most important lines for inclusion in the doll house books.

Results and Discussion

Daily Comprehension Measures

The data to be reported first are the percent correct comprehension questions on the daily assessment passage. As there was no reliable effect due to question type, the data were collapsed across this variable. The daily comprehension scores are shown in Figure 5. Students 1 and 2

INSERT FIGURE 5 ABOUT HERE

---

received four days of baseline while Students 3 and 4 received 6 and Students 5 and 6 received 8 baseline days; in all other respects, the treatment was the same. Visual inspection suggests that the pattern of improvement was similar to that found in Study 1. The six students of Study 2 had baseline accuracy not exceeding 40% correct. They proceeded to make stepwise progression towards means in excess of 70%. Four of the six students reached a stable level of 80% for five successive days, taking 12, 11, 11, and 12 days respectively to do it (Students 1, 3, 4, and 6). Student 5 reached criterion of 70% correct in 12 days. Student 2 was the only "failure"; she progressed from a baseline of 12% correct and reached a steady level of 50% correct in 12 days, a significant improvement, but she never approached the 70-80% criterion level of the remaining five students. All students maintained their improved level of performance on both short- and long-term maintenance.

A series of planned comparisons was conducted on the various phases. Confirming the visual impression, mean accuracy during training was significantly higher than during baseline, $F(1,20) = 243$, $p < .001$. Students continued to gain in accuracy during the second half of treatment and the difference between first half and second half of training was reliable, $F(1,20) = 38.84$, $p < .001$. This level was maintained for there were no significant differences in accuracy between the second half of

training and maintenance or between maintenance and the long-term follow-up that took place eight weeks after the termination of training. The six matched students who did not take part in training did not improve over the four week hiatus between pre- and post-testing. Their baseline and eight week follow-up data are plotted in Figure 6, together with the comparable data from the experimental subjects. Four of the six control students attained lower or equivalent scores when comparing follow-up with baseline performance. Two students showed slight improvement, but neither student ever attained the criteria achieved by the experimental students.

INSERT FIGURE 6 ABOUT HERE

Dialogue Changes

We also examined the quality of the summary statements and questions generated during the reciprocal teaching sessions. A similar pattern to Study 1 emerged, unclear and "generated with assistance questions" declined over the course of training from 25% in the initial third to 15% in the final third. Detail questions remained fairly low and stable and main idea questions increased from 54% to 70% of the total.

The quality of the summary statements also improved with incorrect and incomplete statements declining from 18.3% to 10.3% and detail summaries declining from 29% to 4%. These declines in inadequate summaries were accompanied by a significant increase in main idea summaries ($z = 2.86$, $p < .002$) from 52% to 85% of the total.

In summary, there was a definite improvement in the quality of the students' dialogues during the course of training. At the outset, students required more assistance with the dialogue, asked more unclear and detailed questions, and made more incomplete/incorrect or detailed summaries than they did on the last intervention day. Both main idea questions and paraphrases increased over time.

Students improved at differential rates. For example, Student 6, a minority student whose Slossen test indicated an IQ of 70, made steady but slow progress as indicated by the dialogues shown in Table 5. The data are taken from Days 1 to 15, the day on which he reached criterion.

INSERT TABLE 5 ABOUT HERE

As a further check on the improvement in dialogue, selections of verbatim transcripts of three sessions for each group were rated by two independent raters. The transcripts, from the beginning, middle, and end of intervention, were randomized and the raters' job was to rank them depending on whether they thought the dialogues were from the initial, middle, or final phase of intervention. Percentage of agreement, determined by the number of times the raters correctly identified the order of each transcript, was 83% for initial and final phases and 67% for the transcripts from the middle segment of training.

Generalization probes taken in the classroom setting resulted in variable performance but did show clear evidence of improvement. Probes were taken in two settings, social studies and science. At baseline on the social studies probe the range of percentile rankings was .9 to 43, with four students at or below the fifth percentile. The percentile rankings were typically higher in science with a range of 2-47, four students scoring at or below the 25th percentile. Although performance on these probes was variable, the total mean change in percentile rankings (combined across settings and phases) were: Student 1, 47; Student 2, -.5; Student 3, 26; Student 4, 35.5; Student 5, 40.6; and Student 6, 36. Excluding Student 2, at the conclusion of the study, the range of subjects' mean percentile ranks was 49-76. All students, except Student 2, demonstrated considerable generalization to the classroom setting. Student 2 was also the only student who did not reach criterion during the intervention. Figure 7 illustrates the percentile rank achieved by the experimental

---

INSERT FIGURE 7 ABOUT HERE

---

students in baseline and then four months later in follow-up. To show that the increment and percentile ranking achieved by the participants in Study 2 exceeded the variability in percentile rankings one might ordinarily see over the course of four academic months, it is helpful to compare the performance of six other students who scored at or below the 25th percentile on the baseline measure (see Figure 7). Although these

"control" students typically attained higher percentile rankings during baseline, none of them surpassed the 45th percentile in follow-up.

In addition, a posttest administration of the Gates MacGinitie (Form 1) indicated that four of the six students in Study 2 improved on the comprehension subtest of this measure. The results of the pretest are included in Table 4. The following positive differences were attained between pre- and posttest scores on the comprehension measure: S1, 2 months; S2, 0 months; S4, 1 year 5 months; S5, 1 year 3 months; S6, 1 year 8 months for a mean of 7.3 months. It should be noted that the pretest was administered on a large group basis, while the posttest was administered to the pairs of students. However, since corresponding increments were not observed on the vocabulary subtest (mean gain 1.3 months), we are disinclined to attribute the comprehension gains to the testing situation.

Transfer Tests

Summarization. The first transfer test to be considered is the simplified summarization test. Independent raters scored the students' summary sheets and assigned points as follows: one point for each list of exemplars crossed out, one point for each superordinate given, one point for trivial or redundant ideas which were deleted. In addition, each idea unit in the passages used had been assigned a number indicating its importance (see Brown & Day, in press, for details)--one being unimportant and four being important. If a student in the present investigation underlined a topic sentence rated as a 3, that was worth one point. If a student underlined a topic sentence rated as a 4, that was accorded two

points. The same procedure applied to inventions. The total number of points earned for the passages summarized during the pre-testing and those summarized during post-testing was tallied and used in the analysis.

The mean for the pre-test summary score of students in Study 2 was 36.33 and 46.33 on the post-test; this difference was reliable, $t(5) = 2.44$, $p < .05$. More interesting than the total scores, however, is a profile of the gains. Only small gains were made on the superordination rule (6% crossing out lists and 5% naming lists); however, the students were quite facile at this on the pre-test. They were not as adept initially at selecting or inventing topic sentences and did improve 20% in these abilities. However, the major gains came in the deletion of redundant and trivial material (33%) and in the importance ratings assigned to their topic sentences (36%). The six poor comprehenders who served as a pre-post control group earned 36 points on a pre-test and 34 points on a post-test of the summary task. It would appear that the continual instruction during training to paraphrase prose segments by concentrating on the main idea, did lead to significant transfer to a quite dissimilar task.

Question prediction. The second transfer measure concerned predicting questions. Two independent raters were given typewritten copies of the ten questions per passage generated by the students (corrected for grammar and spelling). They were asked to rate each question as: a main idea question (worth 2 points) or a detail question (1 point), as a question lifted directly from text (0 points) or paraphrased (1 point). In addition, the

quality of each question was rated on a five point scale ranging from 1 (very poor) to 5 (excellent). Finally, a question which the rater indicated she was likely to ask if evaluating the student's understanding of the passage was accorded 1 point. The total number of points earned in each category was tallied. The points were then summed across the two sets of questions generated during pretesting and the two sets produced during posttesting. The raters were trained in the scoring procedure together and jointly evaluated ten question sets until they arrived at better than 95% reliability on the rating of the main idea and paraphrase categories. They then each rated all the remaining questions independently. The Pearson Product Moment correlation coefficient calculated to yield interrater reliability was .88.

The pre-post scores (out of a possible 100) were 44.83 and 62.50; this difference was not reliable, $t(5) = 1.73$, $p > .05$. However, consider the starting level of these students against the level set by the 13 control good comprehenders. The good comprehenders scored 64 points on this task. Prior to training, two of the students in Study 2 had scores of 60 (Student 3) and 66 (Student 4), well in the normal range and they did not improve as a function of training. The remaining four students had much lower entering scores (Student 1 -- 56, Student 2 -- 31, Student 5 -- 55, and Student 6 -- 0) but all students concluded the study earning between 58 and 65 points. The improvement of the four initially poor students was reliable ($p < .05$). The six poor comprehenders not participating in the training achieved 48 points on the pre-test and 52 on the post-test ($p > .05$).

Detecting incongruities. A computer print-out indicated the story lines that each student judged as making sense in the story or not making sense. Reported for each student were the number of responses that were false positive, false negative, true positive, and true negative.

A *t* test was conducted to compare accuracy with detecting incongruous sentences on the pretest versus the posttest. There were a total of 36 lines evaluated by each student during pretesting and during posttesting. Of the 36 lines, three were objectively anomalous. Accuracy on this measure was defined as the number of correct detections of incongruous sentences minus three times the percentage of times a student said "no" when evaluating if the sentence made sense in the story, a correction for guessing. Corrected for guessing, the mean pretest score was 1.27 correct and the mean posttest score was 1.92 correct recognition of incongruous sentences. This improvement was significant $t(5) = 4.79$, $p < .001$.

Not only did the rate of detecting incongruities improve but so also did the quality of verbal responses during probes. When the students indicated a line didn't make sense in the story, the investigator would probe, "Can you tell me why this line doesn't make sense in the story?" During pretesting, very typical responses to the probe were, "It just doesn't read right," "It isn't important," "They need to be more specific," "I just don't like that one," or "I don't know." In contrast, on the posttest, even when the students incorrectly evaluated a line as not making sense, their reasons were generally more specific and informative, e.g., "They said the boat was in the water, so how can it be off the shore?" "The

recipe didn't tell you what they were making," or "Under a slide they usually have sand, not grass." One comment which teachers would appreciate was made several times when students were appraising Day in the Park: "It says, 'the teacher is very tired' but teachers don't get tired," or more enlightening still, "It says, 'the teacher is very tired' but they don't do anything."

Ratings of thematic importance. The data for this transfer test will not be reported in detail. The test was simply a failure. The students rated items essentially at random before and after training. This performance was in sharp contrast to that of the 13 "control" good comprehenders who, replicating the original Brown and Smiley (1977) seventh grade data, tended to rate items in concordance with college students. Agreement was particularly high for the lowest and highest ranked units. In a series of studies we have found that children with diagnosed reading or learning problems find the rating thematic importance task extremely difficult (Smiley, Oakley, Worthen, Campione, & Brown, 1977). Note that the students in this study did improve in their ability to select important topic sentences as indicated in the simplified summarization task. Thus we assume that the difficult Brown and Smiley rating task was an insensitive measure of the emergent ability to concentrate on importance at the expense of trivia, demanding as it does that students simultaneously keep in mind large segments of text and rate each and every one of the idea units in terms of fine degrees of importance.

In summary, the main findings of Study 2 are that students diagnosed as experiencing particular problems with reading comprehension improved considerably as a result of taking part in the reciprocal teaching sessions. All students reached asymptote within 12 days, and for five of the six the level was at 70-80% correct, comparable to accuracy attained by the 13 good compehenders who acted as control subjects. Only Student 2 failed to reach the normal level, but she did improve from 15 to 50% and maintained that level well. Indeed, all of the students maintained their asymptotic level for at least eight weeks.

In addition to this dramatic increase on the daily comprehension measures, the students improved their percentile ranking in the classroom, gaining an average of 37 percentile points. The quantitative improvement in the ability to answer comprehension questions on texts read in a variety of settings was accompanied by a qualitative improvement in the students' dialogues. Main idea statements and summaries came to predominate, and unclear, incomplete or detailed responses dropped out.

There was also encouraging evidence of transfer to new tasks. Reliable improvement was found in the ability to use condensation rules for summarizing, in the ability to predict questions that a teacher might ask concerning a text segment and in the ability to detect incongruous sentences embedded in prose passages.

Given the success of Studies 1 and 2, we decided to attempt another replication, but this time the teacher would be a "real" teacher, not an investigator, and the instruction would take place in naturally occurring

groups within the school setting. In Study 3, we looked at four groups of students, two classroom reading groups for the poorest readers and two reading groups that met regularly in a resouce room. In all other respects the study was a replica of Study 2.

STUDY 3

Method

Subjects. The students were seen in the reading groups they normally attended. Originally, six teachers were contacted and five were willing to participate. After screening the students assigned to these teachers, one group was dropped because the students did not meet the decoding requirements for entry into training. Of the remaining four groups, two were taught by regular classroom teachers (Groups 1 and 4) and two by resource room teachers (Groups 2 and 3), all in rural schools in central Illinois.

The majority of the students were seventh graders. However, Group 4 was an eighth grade grouping, one student in Group 1 was an eighth grader and two students of Group 2 were sixth graders. Seven of the 21 students were female, all were white. The standardized scores for each subject are shown in Table 6. Because the participating schools used different tests,

---

INSERT TABLE 6 ABOUT HERE

---

grade equivalents are shown in Table 6. All students met the decoding

criteria of 80 wpm correct with no more than two errors, when orally reading seventh grade texts. All students scored 50% or less correct on the baseline assessment comprehension passages. Their standardized comprehension scores were variable ranging from approximately at grade level (subject 5 of Group 1 = -.8, subject 4 of Group 2 = -.2, and subject 3 of Group 4 = -.6 years delayed) to several years delayed (subject 6 of Group 1 = -5 years, and subject 4 of Group 4 = -4.1 years delayed). In general, however, the students averaged two year delays in reading comprehension scores (group means = Group 1 = -2.24, Group 2 = -1.98, Group 3 = -1.96 and Group 4 = -2.35).

To summarize, all the subjects were junior high students identified by their schools as requiring supplementary or special reading instruction, but not labeled as LD or EMR. Standardized testing and our own inventory indicated that they were all adequate decoders but poor comprehenders, tyically comprehending at about two years below grade level. The students were all instructed by experienced teachers in their natural groups. They shared educational and environmental backgrounds more alike than dissimilar.

Material. The materials were identical to Study 2. The teachers, although free to progress at their own rate with the training passages, covered the material at approximately the same speed.

Procedure

With the exception that the training was conducted by the teachers, the procedures for Study 3 were identical to those of Study 2. The teachers received three training sessions. In the first, they were introduced to the rationale behind the reciprocal teaching intervention and shown the results of Study 1. They also viewed a video-tape of the first author employing the technique with a group of students.

In the second training sessions, the teacher and the first author practiced the procedures privately, with the first author modeling both the teacher's role and behaviors that might be expected from students. Difficulties that could arise were anticipated and discussed, such as situations where a student is unable to generate a questions, or where a student summarizes by reiterating the whole paragraph in detail! Remedial steps were demonstrated, etc.

In the final session, the teacher and the first author met with a group of seventh graders who were not taking part in the study and practiced the procedure. The first author modeled how the procedure should be introduced to the students, modeled the four main activities, and the process of feedback. The teachers then assumed responsibility for the group and, as the practice session transpired, the teacher and investigator discussed the proceedings with one another. In addition, the teachers were left with several pages of directions regarding the introduction and daily format of the training sessions. The first author also checked weekly on the teacher-directed sessions to see if the intervention was being

conducted properly. These visits provided further opportunity for discussion and resolution of any difficulties encountered. The students were shown their progress charts on a daily basis during baseline, maintenance, and longterm follow-up and on a weekly basis during intervention and their improvement was discussed with them. All reciprocal teaching sessions were tape-recorded.

Results and Discussion

Again no differences due to question type (text-explicit, etc.) were found and, therefore, the data were collapsed across this variable. The four groups of subjects were subjected to different amounts of baseline (from 4-10 days) otherwise they were treated identically. Individually, the students performed in a manner similar to that found in Study 2 (see Palincsar, 1982, for full details). All of the subjects in Groups 1-3 individually reached criterion within 15 days. In Group 4, all students reached criteria in 5 days. If one considers the group means, two groups reached criterion in 13 days (Groups 1 and 2) one in nine days (Group 3) and one in five days (Group 4). It is interesting to note that in Group 4 (the only eighth grade grouping), two of the four students were performing excellently on the first day. The resultant group in some sense consisted of three models, the teacher and the two good students, and two tutees, the remaining two poor students. In this favorable milieu, the poor students rapidly improved, and the entire group reached criterion in 4 days, versus a mean of 12 days for the other groups. Such findings if replicated could have important implications for decisions concerning the composition of the "optimal" reading group.

The daily comprehension means per group are shown in Figure 8. Students were typically achieving 40% accuracy on comprehension

---

INSERT FIGURE 8 ABOUT HERE

---

questions during baseline. With the introduction of the intervention, their accuracy increased steadily, if gradually, until all groups were consistently scoring about 70% by the fifteenth day of intervention. The students continued to show gains during maintenance with slight decrements during follow-up.

To confirm these observations, phase contrasts identical to those conducted for students in Study 2 were conducted for students in Study 3. Mean accuracy during training was significantly better than during baseline, $F(1,80) = 487$, $p < .001$ as was accuracy in the second vs. the first half of training, $F(1,80) = 76.701$, $p < .001$. Unlike Study 2, performance continued to increase during maintenance, $F(1,80) = 5.72$, $p < .02$. Although this difference is reliable it represents only a 3 percentage point increase. There was a slight decline at the eight week follow-up compared with immediate maintenance performance, $F(1,80) = 7.61$, $p < .01$, but again this represented a five point difference in actual scores. Performance on follow-up was equivalent to the last few day of training. In short, students in Study 3 started at a level of approximately 40% correct and ended at a level of 80% correct, an impressive effect of training.

Quality of dialogue. A similar improvement in quality of dialogue was found as in Studies 1 and 2 but was less dramatic in Study 3. In the group settings, the teachers decided to call upon the "better students" in the inital sessions and then gradually to introduce the poorer students into the dialogue as they felt they could handle the responsiblity, a natural procedure for experienced teachers. This resulted in a level of student responses that was higher initially and did not improve as dramatically over sessions. In addition, the training materials were not randomly sequenced. Rather, care was taken to select three training passages with which to begin the intervention which were well organized (used a number of subheadings), contained concrete subject matter, and were composed of relatively brief paragraphs (three to four sentences). It was decided that sequencing the material from easy to hard would facilitate a successful beginning to the training phase. Typically it took six days to work through these first three passages. It might be anticipated that a greater number of unclear questions, questions generated with assistance, and incorrect/incomplete summary statements would have resulted if the training materials had not been sequenced in such a manner. As the intervention progressed, more difficult texts were used in the training sessions; the content became more technical (e.g., the generation and use of solar energy), less familiar (e.g., the founding of the Inca civilization), and less concrete (e.g., myths recounted to explain the formation of volcanoes). The nature of the passages may have contributed to the observation that there was not as clear an increase in main idea question

types for Students in Study 3 as there had been in Study 1. The trend was still the same, however, with incomplete or unclear questions decreasing significantly from 20% to 4% ($z = 3.18$, $p < .001$) and main idea questions increasing (though not significantly) from 57-70% across the sessions. Similarly, main idea summaries increased from 68% to 85% of the total produced by the groups.

To investigate whether the discrete changes in verbal behavior were reflected in overall qualitiative changes in the dialogues, two raters were asked to independently sequence three transcripts from each of the four groups. The transcripts were selected from the beginning, middle, and end of the intervention phase. Percentage of agreement, determined by the number of times the raters correctly identified the order of each transcript, was 87% for transcripts from the initial segment of training, 63% for the middle segment, and 63% for the final portion of training. The raters correctly identified the sequence the majority of times. They were especially accurate at identifying that segment which occurred first.

Transfer tests. The same pattern of transfer results occurred in Study 3 as in Study 2. As the rating thematic importance task was judged inappropriate and failed to produce reasonable behavior in both studies, it will not be discussed here (see Palincsar, 1982, for full details). The remaining three transfer tests resulted in reliable improvements.

Summarization. The difference between pre- and post-tests scores (38.95, and 48.71 respectively) was significant $t(20) = 3.24$, $p < .004$. The largest gain again occurred in the rating importance of topic sentences

(40%). These students also improved 19% on the most difficult rule, inventing a topic sentence. For example, after reading a seven sentence paragraph about two different men who kept tarantulas, one to deal with cockroaches in his kitchen and another to keep robbers away from his jewelry store, one student in Study 3 crossed out the paragraph and wrote, "Spiders can get rid of disturbers." The tarantulas were not referred to as spiders in the passage, and the words "disturbers" or "disturb" did not appear in the text. Unlike the students in Study 2, there was only a modest gain in deletions (9%), but these students showed much higher pretest competency on this measure.

Question prediction. Out of the total possible score of 100, the students scored 53.62 on pretest and 61.24 on posttest, a significant difference, t (20) = 4.58, $p < .001$. The percentage of gain was distributed in the following manner. The greatest percentage of gain points was earned for quality of questions (62%). Raters determined that posttest questions were more clear and complete than pretest questions. Twenty-one percent of the gain was accounted for by the increased likelihood of the students asking similar questions to those proposed by the raters. Thirteen percent of the gain was attributed to asking more main idea than detail questions. The posttest level of 61.24 points is remarkably similar to the posttest level of 62.50 from Study 2 and the 64.0 level of the 13 good comprehenders included as comparison students in Study 2. Again, the improvement in posttest scores after training is modest but reliable, bringing the poor comprehenders up to the "normal" level.

Detecting incongruities. The mean number of incongruous sentences detected (corrected for guessing) rose from 1.26 on the pretest to 1.84 on the posttest, again a reliable difference, t (20) = 5.60, $p < .001$, similar to that found in Study 2.

In summary, very similar results were found in Studies 2 and 3. The effect of the reciprocal teaching intervention was reliable, durable, and transferred to tasks other than the training vehicle. The similarity of the main results across the three experiments is more striking than the differences. Even though the intervention was one-to-one in Study 1, in small groups in Study 2, and in larger, naturally occurring, groups in Study 3, the same pattern of results pertain. Classroom teachers, receiving only limited introduction to the method, were as effective as the investigator in conducting the intervention.

In this light, it is interesting to note that without exception the teachers expressed a degree of skepticism regarding their students' ability to participate in the reciprocal teaching prior to beginning the study. At the conclusion of the study, the teachers were pleased not only with the progress demonstrated by the students in the reciprocal activities as well as their improvement with the comprehension measures, but by other results as well. The teachers observed that general "thinking" skills seemed to improve. The students appeared better able to locate important information and organize their ideas--skills which the teachers regarded as important "study skills." In confirmation of the teachers' observations, students reported that they were using the instructed activities (primarily

summarizing and question predicting) in their content classes. As one student proudly reported to his reading teacher after a triumphant attempt to write a book report using the activities he had learned in the reciprocal teaching training, "Mrs. P, you'll be glad to hear this wasn't all for nothing."

All of the teachers indicated that they would add reciprocal teaching to their instructional repertoire using it with their more capable readers as well as their poor comprehenders. The one eighth grade teacher planned to divide her class of 20 students into four small groups to which she would assign one student who had been trained in reciprocal teaching. This student would function as group leader in a peer tutoring situation. One of the two remedial reading teachers planned to implement the procedure with her younger students, reading the passages orally rather than silently. The teachers concurred that an important facet of the procedure was sharing with students their progress. While the teachers would not plan to administer routine assessments or tape record the session on every occasion, they would do this periodically to demonstrate progress.

The students' responses to post training questionnaires showed that they also responded positively to the procedure--particularly the opportunity to assume the role of teacher. Students, evaluating the procedure, indicated that "finding the good right question" was the most difficult activity and that summarizing was the most helpful activity.

GENERAL DISCUSSION

This series of studies can be regarded as successful for five main reasons: (a) The effect was large and reliable; of the 10 subjects included in Studies 1 and 2, 9 improved to the level set by good comprehenders and all of the subjects in Study 3 met this level. (b) The effect was durable; maintenance probes showed no drop in the level of performance for up to an eight week period (Studies 2 and 3). Although there was a decline after six months (levels dropping from 70-80% to 50-60%), only one session with the reciprocal teaching procedure was sufficient to raise performance back to the short-term maintenance level (Study 1). (c) The effect generalized to the classroom setting: of the 10 students taking part in Studies 1 and 2, nine showed a clear pattern of improvement, averaging a 36 percentile rank increase, thus bringing them up to at least the average level for their age mates. Given the difficulty reported in obtaining generalization of trained skills across setting (Brown & Campione, 1981; Meichenbaum, 1977), this is an impressive finding. (d) Training resulted in reliable transfer to dissimilar tasks; summarizing, predicting questions, and detecting incongruities all improved. Again this is an impressive finding given prior difficulty with obtaining transfer of cognitive skills training (Brown & Campione, 1978, 1981; Brown, Campione, & Day, 1981). (e) Sizable improvements in standardized comprehension scores were recorded for the majority of subjects. (f) The intervention was no less successful in natural group settings conducted by teachers than it was in the laboratory when conducted by the experimenter.

Some reasons why the current intervention may have been more successful at generating maintenance, generalization and transfer than prior studies are that the training was intensive; the subjects were fully informed about the reasons why these activities were important; the subjects were given explicit information concerning the generality of the activities and their range of utility; the subjects were trained in self-regulatory activities including the checking and monitoring of their own comprehension; and the skills themselves were general comprehension-monitoring activities applicable in a wide variety of reading/studying tasks.

We claim that the direct instruction of ubiquitous skills of comprehension-monitoring, coupled with the subjects' understanding of the reasons why these activities are necessary and work, resulted in the impressive performance reported here. In some sense, however, the studies were multiply confounded in that any one of the activities modeled might have been responsible for the improvement. Given the much more limited success of studies where only one activity has been trained, we doubt this; however, component analyses are needed to pinpoint the most economic package that could result in the type of widespread improvement we report. Given the typically limited outcome of cognitive skills training studies (Brown, Campione, & Day, 1981), however, we advocate the procedure of first obtaining a sizable, durable and generalized effect and then conducting the necessary clean-up operations to determine the sub-components that are primarily responsible for the improvement. Such clarifying procedures are currently underway in our laboratory.

Finally, we would like to point out that training studies are not just exercises in cognitive-engineering with immediate applicability to school settings. They are also direct tests of theory involving degrees of experimental manipulation and control in an area where a great deal of data consist of simple one-shot developmental demonstrations. For example, a great deal of developmental research is correlational in nature and there are problems with interpreting such results. To give an example from our own work, in many of our studies we consider the performance of students who do or do not spontaneously adopt an appropriate text processing strategy and this is often the major variable carrying a developmental trend. For example, fifth and seventh graders, who make adequate rough drafts when paraphrasing (Brown, Day, & Jones, in press) or spontaneously underline or take notes of important text elements, etc. (Brown & Smiley, 1978), perform as well as the majority of twelfth graders, whereas twelfth graders who fail to employ these activities look like fifth graders. This pattern suggests that it is the strategy that leads to efficiency, and developmental trends showing improvement with age are created by the increased proportion of strategic subjects. This is a reasonable interpretation but as the data are primarily correlational, the interpretation is not that simple. It could be that the young spontaneous strategy users are the brightest children and would perform better than their peers on any task, and on the particular task in question without the use of strategies. Even partialling out ability factors such as IQ or reading scores does not totally bypass this problem.

The training study is then an important tool for providing convergent evidence of the importance of the strategy under consideration. First the theorist speculates about the underlying processes involved in reading comprehension. Next is the correlational step, students who read well are also found to perform well on the identified underlying processes, while poor readers experience particular difficulty on just these activities (Baker & Brown, in press a,b). Finally, students who are not using the strategy are given training designed to induce the use of processes theoretically specified as key activities underlying efficiency. Others are not. If the theory is correct (and training adequate), and these are the underlying effective processes, trained students' performance should become more like that of spontaneous users. There are nontrivial problems with interpreting the outcomes of training studies (Brown & Campione, 1978, 1981) but they do provide an important manipulative tool to aid theory development. Thus from the point of view of both theory development and successful cognitive engineering, training studies such as these reported here are valuable tools for enhancing our understanding of the mechanism of reading comprehension.

References

André, M. D. A., & Anderson, T. H. The development and evaluation of a self-questioning study technique. Reading Research Quarterly, 1978-79, 14, 605-623.

Baker, L., & Anderson, R. I. Effects of inconsistent information on text processing: Evidence for comprehension monitoring (Tech. Rep. No. 203). Urbana: University of Illinois, Center for the Study of Reading, May 1981.

Baker, L., & Brown, A. L. Metacognition and the reading process. In P. D. Pearson (Ed.), A handbook of reading research. New York: Longman, in press. (a)

Baker, L., & Brown, A. L. Cognitive monitoring in reading. In J. Flood (Ed.), Understanding reading comprehension. Newark, Del.: International Reading Association, in press. (b)

Bereiter, C., & Scardamalia, M. From conversation to composition: The role of instruction in a developmental process. In R. Glaser (Ed.), Advances in instructional psychology (Vol. 2). Hillsdale, N.J.: Erlbaum, 1980.

Brown, A. L. Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), Advances in instructional psychology. Hillsdale, N.J.: Erlbaum, 1978.

Brown, A. L. Metacognitive development and reading. In R. J. Spiro, B. C. Bruce, & W. Brewer (Eds.), Theoretical issues in reading comprehension. Hillsdale, N.J.: Erlbaum, 1980.

Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. Learning, remembering, and understanding. In J. H. Flavell & E. M. Markman (Eds.), Carmichael's manual of child psychology (Vol. 1). New York: Wiley, in press.

Brown, A. L., & Campione, J. C. Permissible inferences from cognitive training studies in developmental research. In W. S. Hall & M. Cole (Eds.), Quarterly Newsletter of the Institute for Comparative Human Behavior, 1978, 2(3), 46-53.

Brown, A. L., & Campione, J. C. Inducing flexible thinking: A problem of access. In M. Friedman, J. P. Das, & N. O'Connor (Eds.), Intelligence and learning. New York: Plenum, 1981.

Brown, A. L., Campione, J. C., & Day, J. D. Learning to learn: On training students to learn from texts. Educational Researcher, 1981, 10, 14-21.

Brown, A. L., & Day, J. D. Macrorules for summarizing texts: The development of expertise. Journal of Verbal Learning and Verbal Behavior, in press.

Brown, A. L., Day, J. D., & Jones, R. S. The development of plans for summarizing texts. Child Development, in press.

Brown, A. L., & Palincsar, A. S. Inducing strategic learning from texts by means of informed, self-control training. Topics in Learning and Learning Disabilities, 1982, 2(1), 1-17.

Brown, A. L., Palincsar, A. S., & Armbruster, B. B. Inducing comprehension-fostering activities in interactive learning situations. In H. Mandl, N. Stein, & T. Trabasso (Eds.), Learning from texts. Hillsdale, N.J.: Erlbaum, in press.

Brown, A. L., & Smiley, S. S. Rating the importance of structural units of prose passages: A problem of metacognitive development. Child Development, 1977, 48, 1-8.

Brown, A. L., & Smiley, S. S. The development of strategies for studying texts. Child Development, 1978, 49, 1076-1088.

Campione, J. C., & Armbruster, B. B. Acquiring information from texts: An analysis of four approaches. In S. Chipman, J. Segal, & R. Glaser (Eds.), Thinking and learning skills: Relating instruction to basic research. Hillsdale, N.J.: Erlbaum, in press.

Chipman, S., Segal, J., & Glaser, R. Cognitive skills and instruction. Hillsdale, N.J.: Erlbaum, in press.

Collins, A., & Smith, E. E. Teaching the process of reading comprehension. In D. K. Detterman & R. J. Sternberg (Eds.), How and how much can intelligence be increased. Norwood, N.J.: Ablex, in press.

Dansereau, D. F. Learning strategy research. Paper presented at NIE-LRDC Conference on Thinking and Learning Skills, University of Pittsburgh, October 8-12, 1980.

Dansereau, D. F., Collins, K. W., McDonald, B. A., Halley, C. D., Garland, J. E., Diekkoff, G., & Evans, S. H. Development and evaluation of an effective learning strategy program. Journal of Educational Psychology, 1979, 71(1), 64-73.

Day, J. D. Training summarization skills: A comparison of teaching methods. Unpublished doctoral dissertation, University of Illinois, 1980.

Frase, L. T., & Schwartz, B. J. Effect of question production and answering on prose recall. Journal of Educational Psychology, 1975, 67, 628-635.

Fry, E. Fry's readability graph: Clarification, validity, and extension to level 17. Journal of Reading, December 1977, 242-252.

Garner, R. Monitoring of understanding: An investigation of good and poor readers' awareness of induced misconception of text. Journal of Reading Behavior, 1980, 12, 55-64.

Harris, P. L., Kruithof, A., Terwogt, M. M., & Visser, P. Children's detection and awareness of textual anomaly. Journal of Experimental Child Psychology, 1981, 31, 212-230.

Kintsch, W., & van Dijk, T. A. Toward a model of text comprehension and production. Psychological Review, 1978, 85, 363-394.

Klahr, D., & Siegler, R. S. The representation of children's knowledge. In H. W. Reese & L. P. Lipsitt (Eds.), Advances in child development and behavior (Vol. 12). New York: Academic Press, 1978.

Larkin, J. H., Heller, J. I., & Greeno, J. G. Instructional implications of research on problem solving. In W. J. McKeachie (Ed.), Cognition, college teaching, and student learning. San Francisco: Jossey-Bass, 1980.

Linden, M., & Wittrock, M. C. The teaching of reading comprehension according to the model of generative learning. Reading Research Quarterly, 1981, 1, 44-57.

Lovitt, T. C., & Hansen, C. L. Round one--placing the child in the right reader. Journal of Learning Disabilities, 1976, 6, 347-353.

Manzo, A. V. Improving reading comprehension through reciprocal questioning. Unpublished doctoral dissertation, Syracuse University, 1968.

Markman, E. M. Realizing that you don't understand: A preliminary investigation. Child Development, 1977, 46, 986-992.

Markman, E. M. Problems of logic and evidence. Behavioral and Brain Sciences, 1978, 1, 194-195.

Markman, E. M. Realizing that you don't understand: Elementary school children's awareness of inconsistencies. Child Development, 1979, 50, 643-655.

Markman, E. M. Comprehension monitoring. In S. Chipman, J. Segal, & R. Glaser (Eds.), Cognitive skills and instruction. Hillsdale, N.J.: Erlbaum, in press.

Meichenbaum, D. Cognitive behavior modification: An integrative approach. New York: Plenum Press, 1977.

Meichenbaum, D., & Asarnow, J. Cognitive behavioral modification and metacognitive development: Implications for the classroom. In P. Kendall & S. Hollon (Eds.), Cognitive behavioral interventions: Theory, research, and procedure. New York: Academic Press, 1978.

Meichenbaum, D., & Goodman, T. Training impulsive children to talk to themselves. Journal of Abnormal Psychology, 1971, 77, 115-126.

Newell, A. One final word. In D. T. Tuma & F. Reif (Eds.), Problem solving and education: Issues in teaching and research. Hillsdale, N.J.: Erlbaum, 1979.

Norman, D. A. Twelve issues for cognitive science. Cognitive Science, 1980, 4, 1-32.

Palincsar, A. S. Improving the reading comprehension of junior high students through the reciprocal teaching of comprehension-monitoring strategies. Unpublished doctoral dissertation, University of Illinois, 1982.

Pearson, P. D., & Johnson, D. D. Teaching reading comprehension. New York: Holt, Rinehart & Winston, 1978.

Resnick, L. B., & Glaser, R. Problem solving and intelligence. In L. B. Resnick (Ed.), The nature of intelligence. Hillsdale, N.J.: Erlbaum, 1976.

Smiley, S. S., Oakley, D. D., Worthen, D., Campione, J. C., & Brown, A. L. Recall of thematically relevant material by adolescent good and poor readers as a function of written versus oral presentation. Journal of Educational Psychology, 1977, 69, 381-387.

Wong, B., & Jones, W. Increasing metacomprehension in L. D. normally-achieving students through self-questioning training. Unpublished manuscript, Simon Fraser University, 1981.

Footnotes

This research was supported by Grants HD06864, HD05951 from the National Institute of Child Health and Human Development and from the National Institute of Education under Contract No. US-NIE-C-400-76-0116. The authors would like to express their appreciation to Kathryn Ransome, Reading Coordinator for the Springfield, Illinois School District, for making Studies 2 and 3 possible. The authors also wish to acknowledge the unflagging cooperation and skillful instruction of the teachers who participated in Study 3: Denise Cananit, Jerry Kirback, Sharon Poynter, and Nancy Richardson.

The major part of Studies 2 and 3 were submitted as part of the requirement for the Ph.D. by the first author, who would like to thank the members of her dissertation committee for their valuable recommendations and support: Coleen Blankenship, Ann Brown, Laura Jordan, Steve Lilly, and David Pearson.

[1]Copies of the stories are available on request.

[2]Copies of the transcripts are available on request.

TABLE 1

Percentile Ranks on Classroom Generalization Probes -- Study 1

	Phase				
Student	B	T1	T2	M2	T3
1	6.5	18	3	37	46
2	.15	33	79	37	37
3	14.5	18	15	8	37
4	.15	33	38	71	15

TABLE 2
Examples of Student-Generated Questions During Reciprocal Teaching

Main Idea Questions

Why don't people live in the desert?
Where are the grasslands of Australia ideal for grazing?
What does the light on the fish do?
What did these people (the Chinese) invent?
Plans are being made to use nuclear power for what?
What are three main problems with all submarines?
Is there just one kind of explosive?
What are one of the things people used explosives for?
What are the Phillipine officials going to do for the people?

Questions Pertaining to Detail

How far south do the maple trees grow?
What color is the guards' uniforms?
How many years did it take to build the Great Wall?
What are chopsticks made out of?
Tell me where the cats hide?
What was the balloon material made of?
What (on the fish) overlaps like shingles on a roof?
How far can flying fish leap?
What is the temperature along the southern shores of Australia?

Questions Requiring Clarification (and Suggested Appropriate Questions Regarding the Same Material and Ideas)

What was uh, some kings were uh, about the kings? (Why is it that kings did not always make the best judges?)
What were some of the people? (What kinds of people can serve on a jury?)
What was the Manaus built for? Wait a minute. What was the Manaus built for, what certain kind of thing? Wait a minute. O.K. What was the Manaus tree built for? (Why was the city of Manaus built?)
What does it keep the ground? (What effect does snow have on the ground?)
What are the Chinese people doing today, like ... What are they doing? (Why are the Chinese people rewriting their alphabet today?)
There's you know, like a few answers in here and one of my questions is, uh, anything that burns and explodes can be fast enough to ... See, they got names in here. O.K.? (Name some explosives.)
In Africa, India, and the Southern Islands where the sun shines what happens to the people? You know, like ...? (Why do people who live in Africa, India and the Southern Islands have dark skin?)

TABLE 3

Examples of Student-Generated Summary Statements
During Reciprocal Teaching

Statements Regarding the Main Idea

It says if a man does his job real good, then he will do better in his next life.
I learned that they have different kinds of Gods, not just Brahman, every family has their own.
It tells us about the two kinds of camels, what they are like and where they live.
My summary is that the part of the earth that we live on and see and know is the top layer, the crust.
This paragraph talks about what happens when people perspire or sweat. They lose a large amount of salt and they get weakness.

Statements Regarding Detail

It is a pair of fins which look like legs.
The sea horse always swims head up.
There were large lizzards and four eyed fish and 30 foot dandelion.
What I learned is that a submarine went around the world in 84 days.
I learned that Cousteau's first artificial island was in the North Sea.
Professor Charles went 27 miles and rose 2,000 feet in his balloon.
They (the aborigines) don't wear much clothes on.
They (Egyptians) made bread a long time ago.

Statements Which Are Incomplete

They talk about it was the richest island; but it didn't have something, o.k., it was the richest island but didn't have everything. They didn't have something. (Although this was a very rich land, no people lved there.)
If you pick a cherry branch in the winter you will have luck hoping they will bloom early. (If you pick a cherry branch in the winter, you will have no luck with it blooming.)
And uranium can be making explosion that equals a skyscraper. (A small amount of uranium can cause an explosion as great as a skyscraper full of dynamite.)

Examples of Student-Generated Critical/Evaluative Comments

"Boy, the paragraph sure is a mess. It is all over the place."
"I don't see how they can say 'heat lightning occurs on hot summer days.' How could you see it?"
"It says here 'cloud to cloud' then 'cloud to earth.' Wouldn't that be the same thing?"
"The word 'meter' throws me off in this sentence."
"What's the difference between soap and detergent anyway?"
"At first I didn't get this because I thought the word 'pumping' was 'bumping.'"
"I don't know what 'omitting' is."
"I have one, what do they mean by 'far away dreams?'"

TABLE 4

Descriptive Statistics of the Seventh Grade Students in Experiment 2

Subject	IQ[1]	Grade Equivalent[2] Vocab.	Comp.	Baseline Comp. % Correct	Oral Reading wpm Correct	Incorrect
1	88	5.5	4.2	30	99.9	1.7
2	67	4.2	3.6	20	110.0	1.8
3	99	6.5	5.3	20	134.0	0.8
4	92	4.9	6.4	40	131.0	0.7
5	80	5.2	3.9	0	99.0	1.1
6	70	5.2	3.1	30	121.0	0.7
$\bar{X}$	82.6	5.25	4.4	23.3	115.8	1.13

[1]Slosson Intelligence Test
[2]Gates McGinitie Test

TABLE 5

The Acquisition of Question-Asking by One Seventh Grade Student

Day 1:

S: What is found in the southeastern snakes, also the copperhead, rattlesnakes, vipers - they have. I'm not doing this right.

T: All right. Do you want to know about the pit vipers?

S: Yeah.

T: What would be a good question about the pit vipers that starts with the word "why?"

S: (No response)

T: How about, "Why are the snakes called pit vipers?"

S: Why do they want to know that they are called pit vipers?

T: Try it again.

S: Why do they, pit vipers in a pit?

T: How about, "Why do they call the snakes pit vipers?"

S: Why do they call the snakes pit vipers?

T: There you go! Good for you.

Day 4:

S: (No question)

T: What's this paragraph about?

S: Spinner's mate. How do spinner's mate ...

T: That's good. Keep going.

S: How do spinner's mate is smaller than ... How am I going to say that?

T: Take your time with it. You want to ask a question about spinner's mate and what he does, beginning with the word "how."

S: How do they spend most of his time sitting?

T: You're very close. The question would be, "How does spinner's mate spend most of his time?" Now you ask it.

S: How does spinner's mate spend most of his time?

TABLE 5 (continued)

Day 7:

S: How does the pressure from below push the mass of hot rock against the opening? Is that it?

T: Not quite. Start your question with "What happens when?"

S: What happens when the pressure from below pushes the mass of hot rock against the opening?

T: Good for you! Good job.

Day 11:

S: What is the most interesting of the insect eating plants, and where do the plants live at?

T: Two excellent questions! They are both clear and important questions. Ask us one at a time now.

Day 15:

S: Why do scientists come to the south pole to study?

T: Excellent question! That is what this paragraph is all about.

TABLE 6

Descriptive Statistics for Students in Study 3

Group 1					Group 2				
Grade Equivalent		Baseline Comp.	Oral Reading wpm		Grade Equivalent		Baseline Comp.	Oral Reading wpm	
Vocab.	Comp.	% Correct	Correct	Incorrect	Vocab.	Comp.	% Correct	Correct	Incorrect
5.4	6.0	20	80	.75	6.1	5.2	30	113	2
4.9	3.9	30	85	1.5	*5.4	3.8	30	145	.8
5.4	4.2	50	81	1.8	3.6	4.5	20	80	1.8
*7.2	5.7	50	87	.8	*4.5	5.8	30	109	2
6.2	6.2	50	98	1.2	5.8	3.8	20	81	.9
4.9	2.0	50	84	1.9					
6.5	4.0	50	97	.9					
$\bar{X}$ 5.7	4.7	43	87	1.3	5.1	4.62	26	106	1.5

Group 3					Group 4				
Grade Equivalent		Baseline Comp.	Oral Reading wpm		Grade Equivalent		Baseline Comp.	Oral Reading wpm	
Vocab.	Comp.	% Correct	Correct	Incorrect	Vocab.	Comp.	% Correct	Correct	Incorrect
5.1	4.6	50	113	2	*7.0	4.7	30	113	2
5.9	4.6	50	82	1.4	*7.5	6.6	30	129	.7
6.6	5.3	30	100	1.1	*7.6	7.4	20	119	2
7.1	4.7	10	88	2	*6.5	3.9	30	129	.7
4.9	6.0	50	136	1.6					
$\bar{X}$ 5.9	5.0	38	104	1.6	7.2	5.6	28	122	1.4

*All subjects except those marked with * were seventh graders. Subject 4 of Group 1 was an eighth grader as were all the subjects of Group 4. Subjects 2 and 4 of Group 2 were sixth graders.

Figure Captions

Figure 1. Percent correct on the daily assessment passages as a function of intervention type and order.

Figure 2. Long-term maintenance of the effect of the reciprocal teaching intervention.

Figure 3. Changes in the quality of question types during the reciprocal teaching intervention.

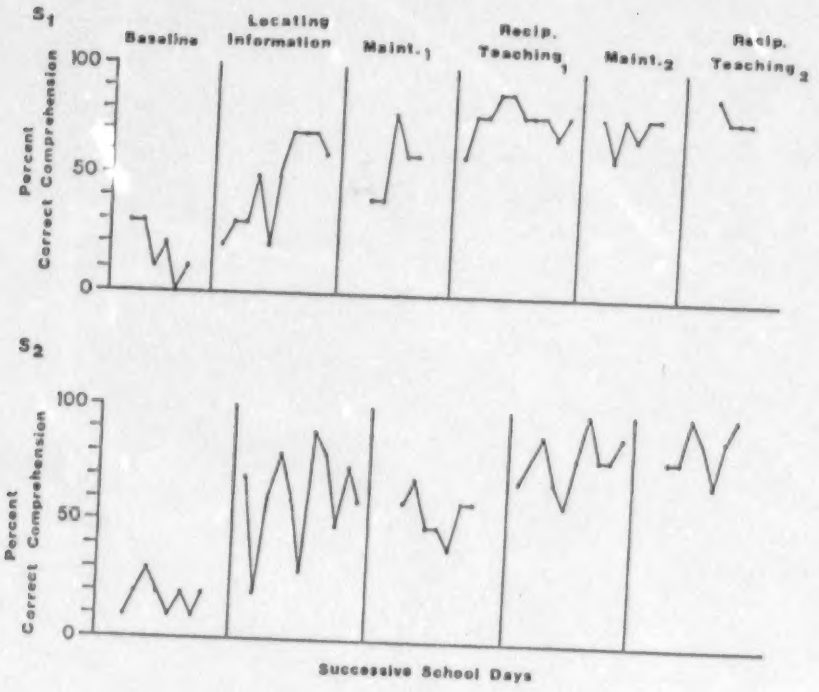
Figure 4. Changes in the quality of the summary statements during the reciprocal teaching intervention.

Figure 5. Percent correct on the daily assessment passages for the experimental subjects of Study 2.

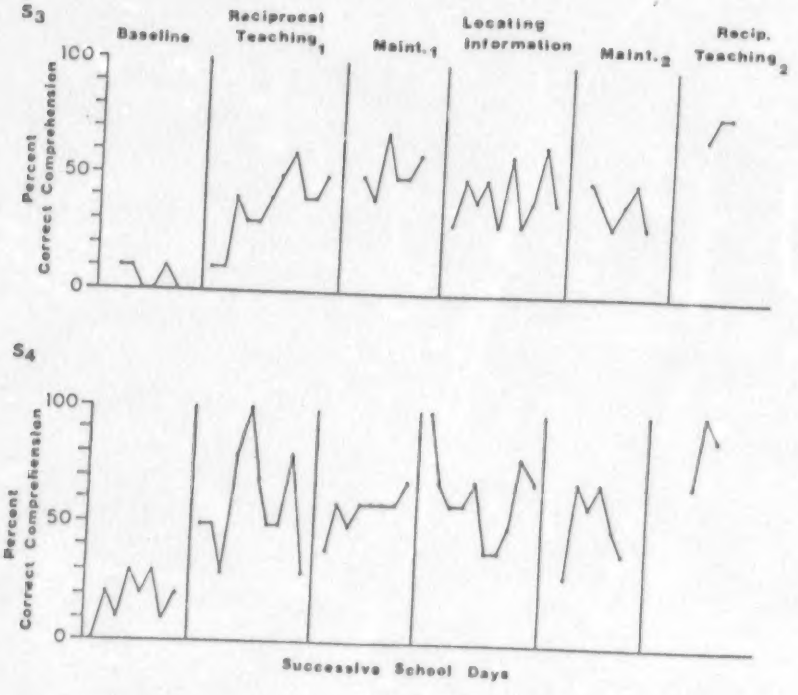
Figure 6. Mean percent correct on baseline and eight week follow-up for the experimental and control subjects of Study 2.

Figure 7. Classroom generalization problems during baseline and follow-up for the experimental and control subjects of Study 2.

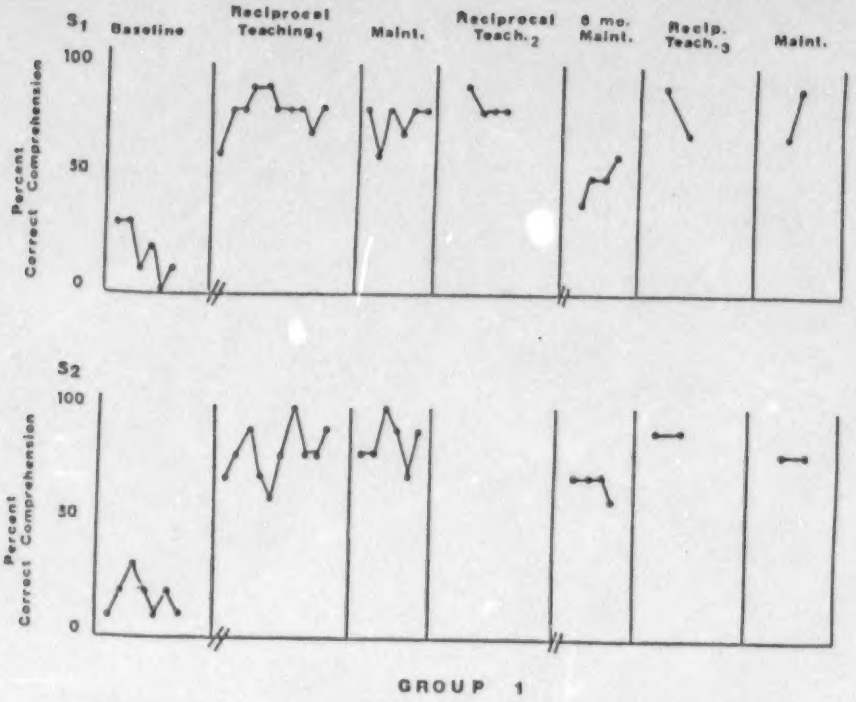
Figure 8. Group mean percent correct on the daily assessment passage for the subjects of Study 3.



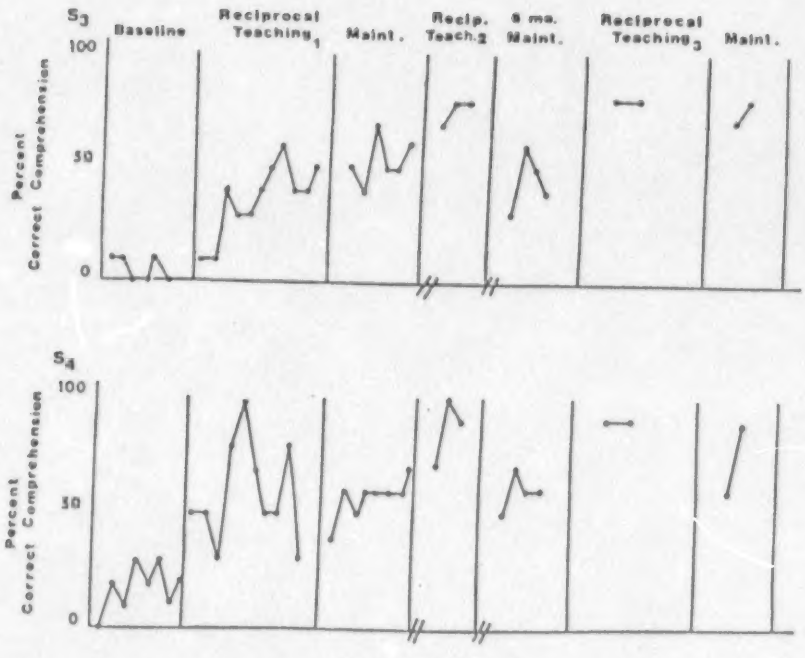
GROUP 1



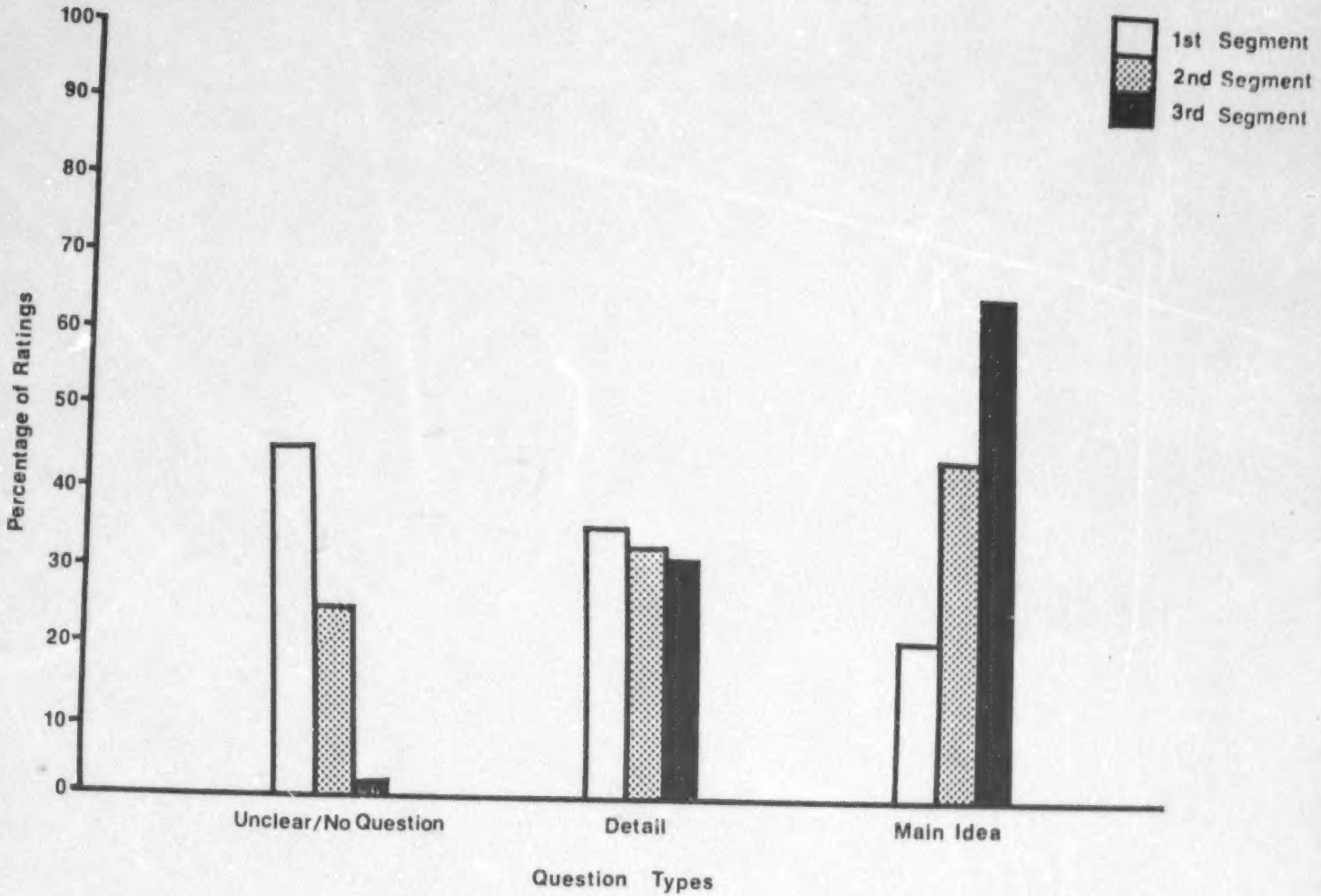
GROUP 2

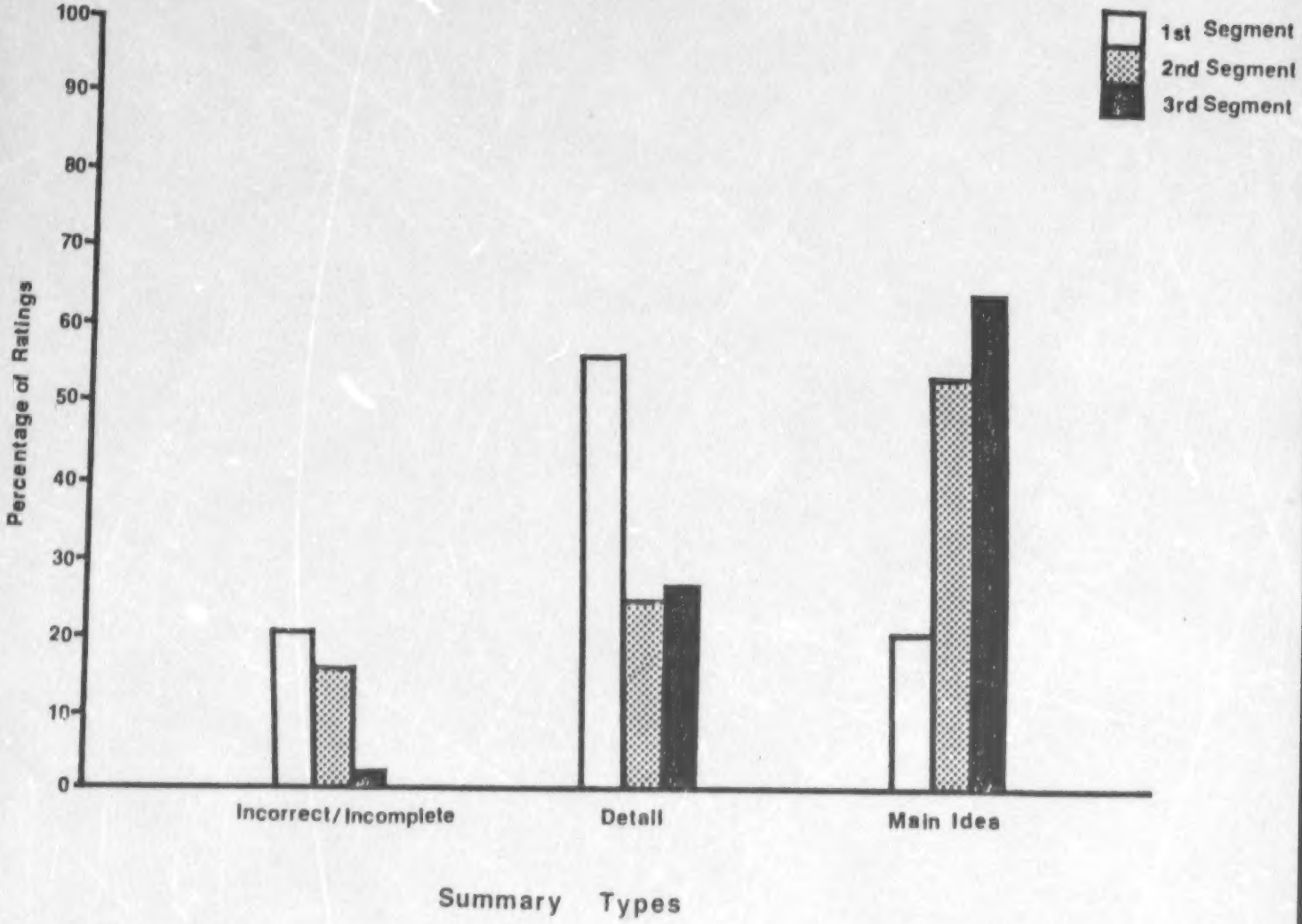


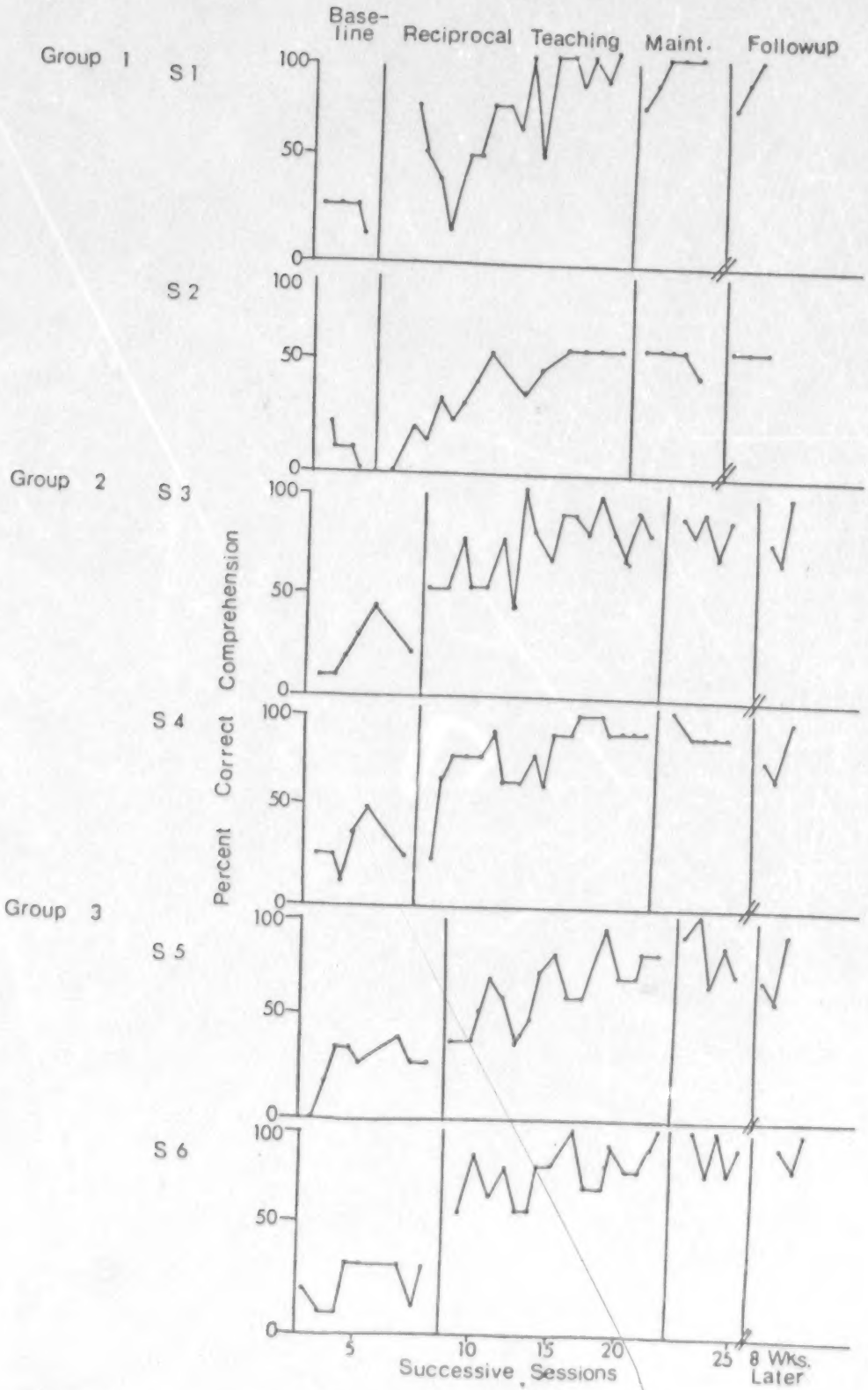
GROUP 1

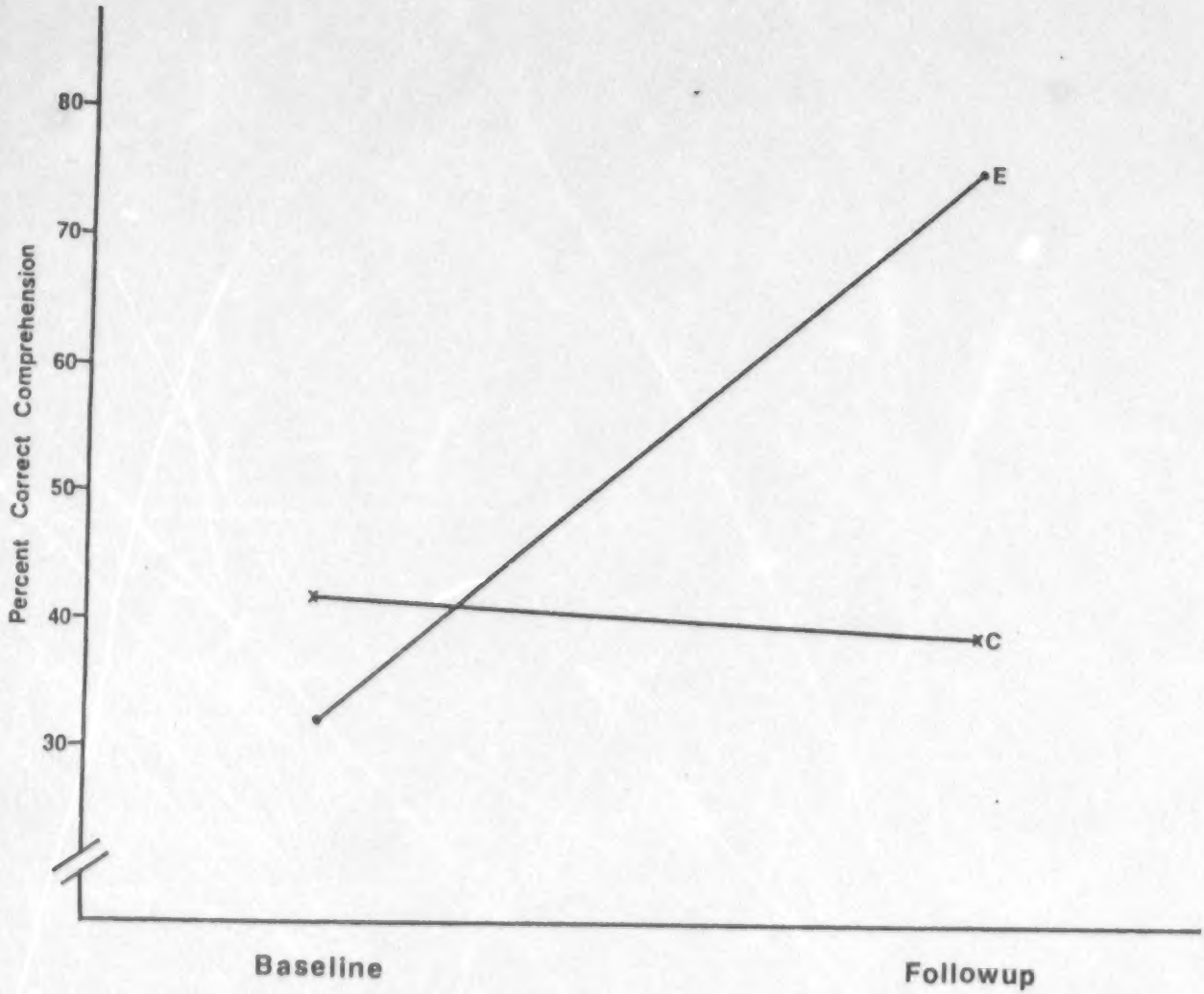


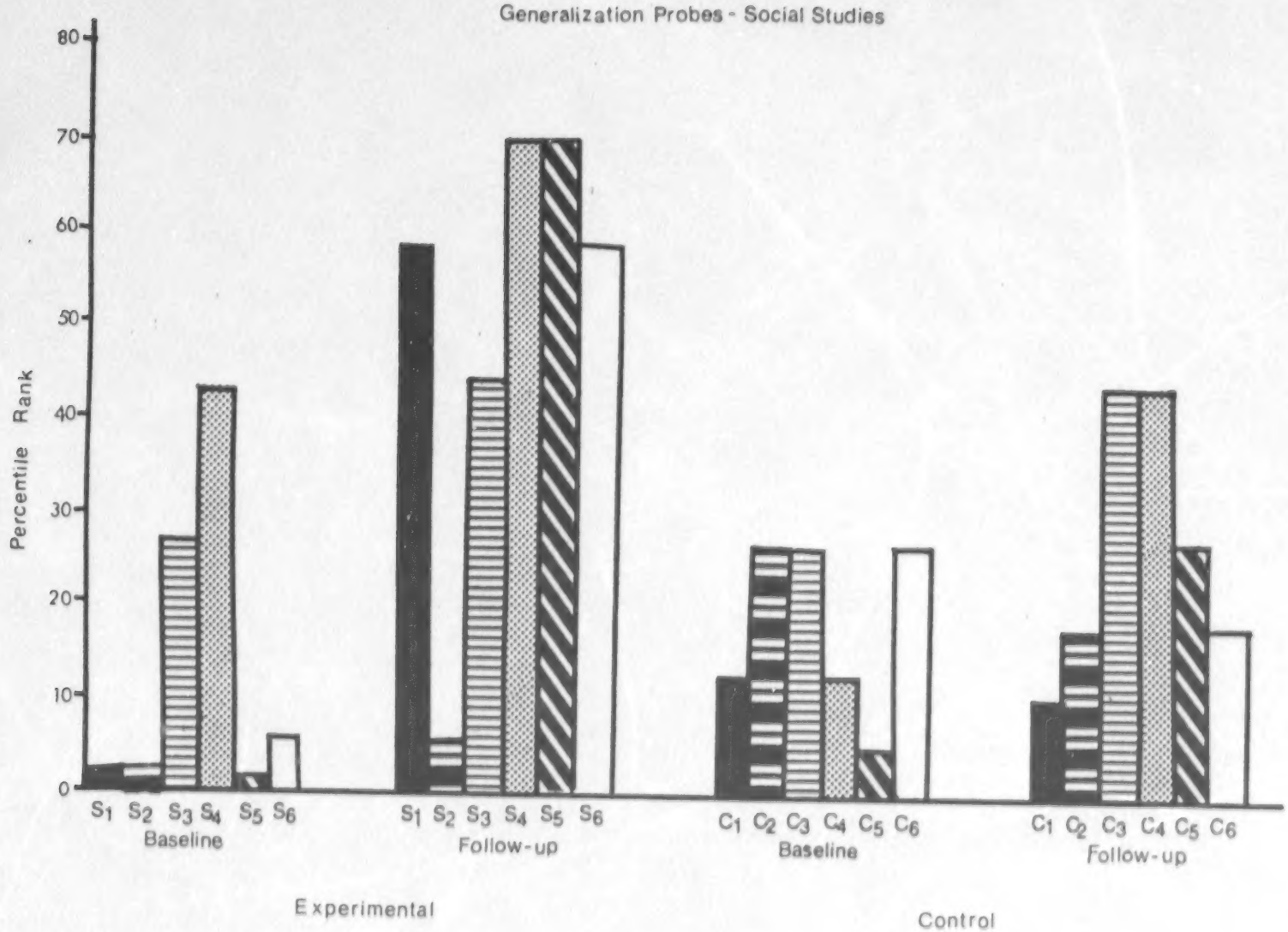
GROUP 2

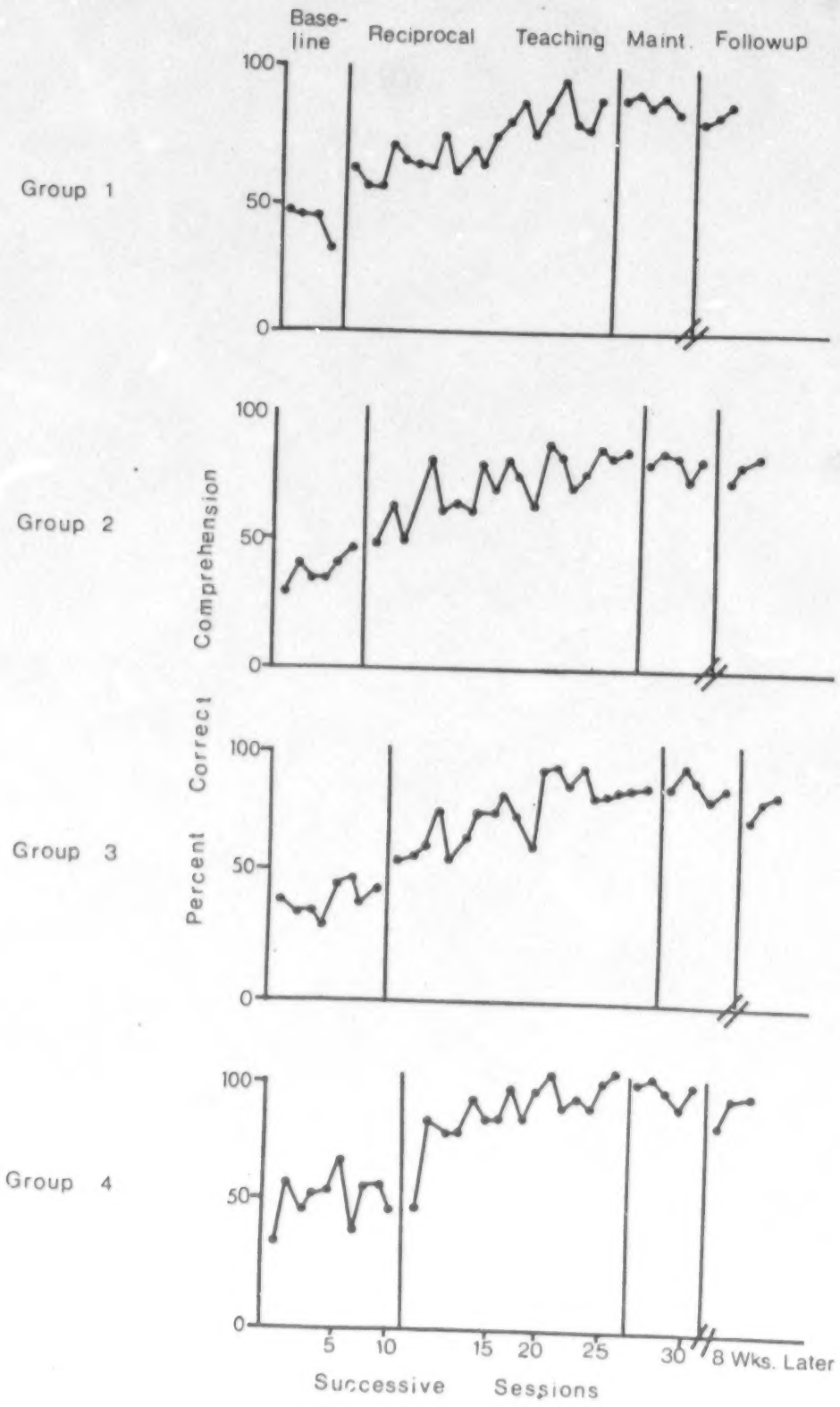

100
90
80
70
60
50
40
30
20
10
0
Percentage of Ratings
1st Segment
2nd Segment
3rd Segment
Unclear/No Question
Detail
Main Idea
Question Types


100
90
80
70
60
50
40
30
20
10
0
Percentage of Ratings
1st Segment
2nd Segment
3rd Segment
Incorrect/Incomplete
Detail
Main Idea
Summary Types


Base-
line
Reciprocal Teaching
Maint.
Followup
Group 1
S 1
S 2
Group 2
S 3
S 4
Group 3
S 5
S 6
100
50
0
Percent Correct Comprehension
5
10
15
20
25
8 Wks.
Later
Successive Sessions


80
70
60
50
40
30
Percent Correct Comprehension
E
C
Baseline
Followup


Generalization Probes - Social Studies
Percentile Rank
80
70
60
50
40
30
20
10
0
S1 S2 S3 S4 S5 S6
Baseline
S1 S2 S3 S4 S5 S6
Follow-up
C1 C2 C3 C4 C5 C6
Baseline
C1 C2 C3 C4 C5 C6
Follow-up
Experimental
Control


Base-
line
Reciprocal
Teaching
Maint.
Followup
Group 1
Group 2
Group 3
Group 4
Percent Correct Comprehension
100
50
0
5
10
15
20
25
30
8 Wks. Later
Successive Sessions